University of Zurich[UZH]

Absolute pitch and relative pitch processing in the human brain

---

Thesis (cumulative thesis)

presented to the Faculty of Arts and Social Sciences

of the University of Zurich

for the degree of Doctor of Philosophy

by

Simon Leipold

Accepted in the fall semester 2019

on the recommendation of the doctoral committee:

Prof. Dr. Lutz Jäncke (main supervisor)

Prof. Dr. Sascha Frühholz

Zurich, 2019

Acknowledgment

I would like to express my great appreciation to Professor Lutz Jäncke for his proficient mentoring and supervision of my doctoral studies. He has fundamentally shaped my academic and personal development from an eager-to-learn master's student to an even more curious scientist. It has been a genuine pleasure for me to study and work under his guidance. I hope this thesis does not merely represent the completion of my doctoral studies, but also the starting point for further successful and productive collaborations between us.

I also thank Professor Sascha Frühholz for his willingness to join my doctoral committee and his support in advancing my academic career.

My sincere gratitude goes to all of my fantastic, current and former colleagues and friends at the Division of Neuropsychology at the University of Zurich. I will never forget my time with you. In particular, I would like to thank Marielle Greber for her incredibly competent and warmhearted support in every imaginable aspect of my academic and non-academic life. I could not have wished for a more brilliant office companion. Furthermore, I would like to thank Carina Klein, not only for being an exceptional co-examiner during my doctoral defense, but also for being a great collaborator and gym buddy; Christian Brauchli for his great ideas and out-of-the-box thinking concerning scientific issues and for always reminding me to never go easy on the gym weights; Silvano Sele for our eternal discussions on the more important (i.e. statistics) and less important (i.e. everything but statistics) aspects of science and for all the great times outside of the office; Lisa-Katrin Kaufmann for her wonderful advice during the preparation of this thesis, and for critical reading and tremendously helpful feedback on an earlier version; Chantal Oderbolz for her excellent contributions to the first published paper of this thesis and for countless hours of proofreading of my manuscripts; Anja Burkhard for her continued support since the start of my master's thesis; and Michi Boos, Matthias Kobi, and Marionna Münger for making the later part of my doctoral studies not feel like work at all.

Last but not least, I am extremely grateful to my family, without whom I would not have achieved anything in life.

This work was supported by the Swiss National Science Foundation (SNSF), grant no. 320030_163149 to Lutz Jäncke.

Abstract

Pitch is one of the main perceptual features of sounds. It is of high importance in music and speech perception and production, and in segregating and grouping the auditory environment into objects. Not all humans perceive pitch equally. Musicians with absolute pitch (AP) preferentially perceive pitches in isolation, whereas all other individuals primarily perceive the relations between pitches. This doctoral thesis compared pitch processing in musicians with AP – the ability to recognize pitches without the help of a reference pitch – to pitch processing in musicians with relative pitch (RP) – the ability to recognize intervals between pitches.

In two separate experiments using electroencephalography (EEG) and functional magnetic resonance imaging (fMRI), a large sample of AP and RP musicians performed a pitch processing task with a listening and a labeling condition. The findings from these experiments are detailed in the three articles of this thesis: *Article 1* describes an attempted replication of an influential EEG study on the electrophysiological correlates of AP and RP. We did not replicate the presence of an AP-specific event-related potential (ERP) component during simple tone listening. However, we partially replicated the reported ERP correlates of RP. The findings from article 1 underline the importance of replication studies in assessing the robustness of extraordinary findings. In *article 2*, we reanalyzed the EEG data using a "brain decoding" framework. Both theoretical and empirical accounts of AP predict a high similarity of neural representations underlying listening and labeling in AP musicians. We confirmed this prediction by demonstrating a higher representational similarity of these task conditions in AP musicians during late stages of pitch processing. *Article 3* describes an fMRI experiment, in which we found lower activation in AP musicians during labeling in the inferior frontal gyrus and the presupplementary motor area, brain regions associated with higher-order cognitive functions. In the listening condition, AP musicians showed higher activation in the planum temporale, a brain region which has been repeatedly associated with AP processing.

The findings of this doctoral thesis contribute to a better understanding of the neural bases of AP and RP processing. They highlight the need for research using large samples sizes, which can be conducted more efficiently in collaboration between multiple research groups. Future work on AP and RP processing will benefit from integrating findings with neural models of auditory processing.

Zusammenfassung

Die Tonhöhe ist eine zentrale Dimension in der Wahrnehmung von Klängen. Sie spielt eine wichtige Rolle bei der Verarbeitung und Produktion von Sprache und Musik, wie auch in der Trennung und Gruppierung von unterschiedlichen Schallquellen in der akustischen Umgebung. Die Tonhöhe wird nicht von allen Menschen gleich wahrgenommen. Musiker mit einem absoluten Gehör nehmen Tonhöhen vorzugsweise isoliert wahr, wobei alle anderen Menschen vorwiegend das Verhältnis zwischen Tönen wahrnehmen. In dieser Dissertation wird die Verarbeitung von Tonhöhen verglichen zwischen Musikern mit absolutem Gehör, der Fähigkeit zur Bestimmung einer Tonhöhe ohne Referenzton, und Musikern mit relativem Gehör, der Fähigkeit zur Bestimmung von Intervallen zwischen Tonhöhen.

Wir untersuchten eine grosse Stichprobe an Musikern mit absolutem oder relativem Gehör während zweier Experimente zur Tonhöhenverarbeitung mittels Elektroenzephalographie (EEG) und funktioneller Magnetresonanztomographie (fMRT). Die Befunde dieser Experimente werden in den drei Artikeln dieser Dissertation beschrieben: *Artikel 1* beschreibt die versuchte Replikation einer wichtigen Studie zu den elektrophysiologischen Korrelaten des absoluten und relativen Gehörs. Wir konnten kein spezifisches ereigniskorreliertes Potential (EKP) für das absolute Gehör nachweisen, aber wir fanden ähnliche EKP-Korrelate des relativen Gehörs wie in der Originalstudie. Diese Ergebnisse zeigen, dass Replikationen wichtig sind, um die Robustheit von aussergewöhnlichen Befunden zu prüfen. In *Artikel 2* haben wir die EEG-Daten unter dem Gesichtspunkt der «Gehirn-Dekodierung» reanalysiert. Wir testeten eine Hypothese, die von theoretischen und empirischen Erkenntnissen zum absoluten Gehör abgeleitet wurde, nämlich dass Musiker mit absolutem Gehör eine ähnlichere neuronale Repräsentation von Tonhöhen-Hören und Tonhöhen-Benennen haben als Musiker mit relativem Gehör. Wir bestätigten diese Hypothese für spätere Phasen der neuronalen Verarbeitung von Tonhöhen. *Artikel 3* beschreibt das fMRT-Experiment. Wir fanden geringere Aktivierung bei Musikern mit absolutem Gehör während des Tonhöhen-Benennens im inferioren frontalen Gyrus und im präsupplementären Motor-Areal. Dies sind Gehirnregionen, die mit höheren kognitiven Funktionen in Verbindung stehen. Während des Tonhöhen-Hörens beobachteten wir bei Absoluthörenden höhere Aktivierung im rechten Planum temporale, einer Gehirnregion, die schon oft mit absolutem Gehör assoziiert wurde.

Die Befunde dieser Dissertation tragen zu einem besseren Verständnis der neuronalen Grundlage des absoluten und des relativen Gehörs bei. Sie zeigen ausserdem den Bedarf an Forschungsprojekten mit grossen Stichproben, welche im Zuge von Kollaborationen mehrerer Forschungsteams effizienter durchgeführt werden können. Zukünftige Forschung zum absoluten und relativen Gehör würde davon profitieren, Befunde im Rahmen von neuronalen Modellen zur auditorischen Verarbeitung einzubetten.

Contents

Chapter 1 General Introduction

Many of the most famous music composers and musicians in history, e.g., Mozart, Beethoven, or Bach, reportedly possessed absolute pitch (AP), the ability to recognize the pitch of a tone as easily as most people recognize the color of an object. Because of this perceived association with extraordinary musical abilities and because of the apparent rarity of the phenomenon, AP is regarded as a mysterious and fascinating ability, of interest for both the general public and academics from diverse fields of research (Deutsch, 2002, 2013). Although it is true that AP is almost exclusively present in musicians (owing to its musical definition), in contrast to the general opinion, AP seems to be a highly specific ability without additional advantages in music or other domains (Zatorre, 2003). In laymen's terms, AP is often called "perfect pitch", but this is misleading as AP musicians typically do not show perfect tone-naming performance and do not possess extraordinary acuity in pitch processing (see e.g., Levitin & Rogers, 2005). However, the very fact that AP is an isolated ability makes it a very attractive target of research for cognitive neuroscientists because it helps to understand how highly specific perceptual and cognitive functions are represented in the human brain.

Musicians with AP are the main subject of study of this thesis. Using both electroencephalography (EEG) and functional magnetic resonance imaging (fMRI), AP musicians were compared to musicians without AP in terms of their neural processing of pitches. The thesis consists of three parts: In the first part (Chapter 1), the reader will be given an introduction to the nature of AP with its many facets. The literature on the neural underpinnings of AP is omitted from Chapter 1 as these studies are reviewed in detail in the empirical part of the thesis. The second part (Chapter 2, Chapter 3, and Chapter 4) comprises the three articles published in the context of this thesis. The last part (Chapter 5) discusses and integrates the findings from the empirical studies and gives an outlook for future research.

1.1 Pitch

A myriad of sounds that humans encounter in their daily lives, e.g., speech, music, and many environmental sounds, have acoustic waveforms that repeat over time. The repetition rate of these sounds is perceived as pitch. Importantly, pitch does not refer to the physical property of a sound itself (i.e. the repetition rate), but rather to the perceptual correlate of this physical property (Oxenham, 2012, 2018; Plack, Barker, & Hall, 2014; Plack, Oxenham, Fay, & Popper, 2005). Pitch is of fundamental importance in music, where it defines melodies, chords, harmonies, and keys. It is also crucial in speech, where it conveys prosodic information in non-

tone languages (e.g., German, English) and additionally semantic information in tone languages (e.g., Mandarin). Finally, pitch contributes to the segregation and grouping of sounds into auditory objects (Patel, 2007; Plack et al., 2005).

The investigation of pitch processing in the brain has a distinguished research tradition, which has been equally informed by research on animals and human participants (Griffiths & Hall, 2012; X. Wang & Walker, 2012). One line of research tries to identify neuronal populations which are specialized for the processing of pitch by contrasting the neural responses to pitch-evoking stimuli with the responses to non-evoking stimuli (Bendor, 2012). Another line of research investigates interindividual differences in pitch perception (Drayna, Manichaikul, de Lange, Snieder, & Spector, 2001; Schneider et al., 2005). For example, individuals with congenital amusia (tone deafness) show specific deficits in pitch processing, and these deficits are related to alterations of neural connectivity between auditory and frontal cortices (Albouy et al., 2013; Peretz, 2016). In contrast, highly trained musicians show enhanced neural responses compared to non-musicians during pitch processing, which are in turn related to neuroanatomical differences in the auditory cortex (Kraus & Chandrasekaran, 2010; Schneider et al., 2002). And then, there exists the special group of musicians with AP that is able to effortlessly recognize, name, and/or produce pitches without the help of a reference pitch (Deutsch, 2013; Ward, 1999; Ward & Burns, 1982). In the following paragraphs, the ability of these musicians is introduced in more detail.

1.2 Absolute Pitch (AP)

The majority of humans perceive pitches in relation to other pitches using an ability called relative pitch (RP). The ability to identify pitch contour by making higher-lower judgements is common to all humans (apart from individuals with amusia) and observable at a very young age (Plantinga & Trainor, 2005). Highly trained musicians are additionally able to identify the exact interval between pitches (McDermott & Oxenham, 2008). In contrast, AP musicians preferentially perceive pitches in absolute terms (Miyazaki & Rakowski, 2002). This preference makes it possible to study AP versus RP processing by comparing AP with RP musicians during the same task.

1.2.1 Historical Aspects

The nature of AP has intrigued musicologists and psychologists for a long time and more recently, researchers from genetics and cognitive neuroscience have joined them in investigating the genesis and the neural underpinnings of the phenomenon (Zatorre, 2003). The term AP (in its German translation *Absolutes Gehör*) was first coined more than a century ago

by the experimental psychologist (and possessor of AP) Karl Stumpf in his work *Tonpsychologie*, where he described case studies of himself and other AP musicians (as cited in Steblin, 1987). Some of the first group studies on AP have been performed by A. Bachem, also an AP possessor, who described different types of AP that he identified by studying more than 100 participants. He concluded that apart from musicians with *genuine* AP which effortlessly identify pitches without a reference pitch, some musicians possess *quasi*-AP and identify pitches using an internal reference pitch in combination with RP. According to Bachem, *quasi-AP* musicians calculate the interval between the internal reference (e.g., the pitch of the tuning standard A4) and the currently perceived pitch (Bachem, 1937).

1.2.2 Prevalence

A review by A. Bachem is also the source of the most commonly cited prevalence estimation of AP, namely that it occurs in less than 1 in 10,000 people of the general population (Bachem, 1955). However, the reported figure is only a rough estimation based on unpublished observations and calculations, and thus the validity of this estimate seems questionable at best (Miyazaki, Makomaska, & Rakowski, 2012). Estimating the prevalence of AP in the general population is a difficult task because different prevalence studies use different methodologies and criteria for establishing AP, and most of the studies sample from a highly restricted population of interest, e.g., music students from a single or a small number of countries (Chavarria-Soley, 2016; Deutsch, Henthorn, Marvin, & Xu, 2006; Leite, Mota-Rolim, & Queiroz, 2016; Miyazaki et al., 2012). The most recent prevalence study included participants from five countries with different cultural and genetic backgrounds but was still restricted to music students. The prevalence estimations using a restrictive criterion ranged from 0-12% among music students in Western countries to 8-57% in East Asian countries (Miyazaki et al., 2018). A large-scale prevalence study sampling from the general population is still missing.

1.2.3 Origins

During the early phase of research into the genetics and neuroscience of AP, it was proposed that AP represents an ideal model system for investigating the joint influences of genetic and environmental factors on the development of a highly specific ability and its neural underpinnings (Zatorre, 2003). Consequently, much research has been devoted to elucidating the necessary or sufficient genetic and environmental factors influencing the genesis of AP (for a comprehensive review see Deutsch, 2013).

From the very start of scientific inquiry into AP, it has been suspected that heredity plays an important role for the development of AP (Ward, 1999). A relatively large number of studies

has reported that AP often aggregates in families suggesting an important role for genetic factors (Bachem, 1940, 1955; Baharloo, Johnston, Service, Gitschier, & Freimer, 1998; Baharloo, Service, Risch, Gitschier, & Freimer, 2000; Gregersen, Kowalsky, Kohn, & Marvin, 1999, 2001; Profita & Bidder, 1988; Theusch & Gitschier, 2011). Further evidence for a genetic influence comes from the observation that AP is usually detected in early childhood when children still have little musical training. Additionally, many AP musicians report that they have known about their ability for as long as they can remember (Deutsch, 2013; Profita & Bidder, 1988). Despite these clues, discovering the genetic basis of the phenomenon is a huge challenge because AP is likely genetically heterogeneous, sample sizes are naturally limited due to the rarity of the phenomenon, and although some first genome-wide studies have been conducted, to date, our knowledge about potential genes involved is still extremely limited (Gregersen et al., 2013; Theusch, Basu, & Gitschier, 2009).

Apart from genetic factors, several environmental factors have been identified to facilitate the development of AP. The most consistently reported factor is age of onset of musical training: Individuals who started training early in childhood have a higher prevalence of AP than individuals who started later (e.g., Bachem, 1940; Baharloo et al., 1998, 2000; Deutsch, Dooley, Henthorn, & Head, 2009; Deutsch et al., 2006; Gregersen et al., 1999). These findings have led to the discussion about a critical period for AP acquisition analogous to the critical period in language acquisition (Deutsch, 2002). The age when this critical period ends has not yet been satisfactorily determined. Attempts to learn AP in adulthood through intensive training have been largely unsuccessful, and until recently the scientific consensus had been that acquiring AP after childhood is extremely difficult if not impossible (Deutsch, 2013). However, two recent preprints (manuscripts posted on the internet that have not yet been peer-reviewed) revisited this topic, and both concluded that acquiring AP in adulthood might not be as impossible as previously thought (Van Hedger, Heald, & Nusbaum, 2018a; Wong, Lui, Yip, & Wong, 2019). Thus, the discussion on acquiring AP in adulthood seems far from over.

A further environmental factor influencing the development of AP concerns the cultural background of the individuals. As mentioned above, AP prevalence is generally higher in East Asian music students than in Western students (Gregersen et al., 1999, 2001; Henthorn & Deutsch, 2007; Miyazaki et al., 2018). This is partly because a certain type of musical training, the so-called *fixed do* training, is more prevalent in East Asian than in Western countries. In *fixed do*, solfège syllables (*do*, *re*, *mi*, *fa*, etc.) are associated with specific pitches (e.g., “do” always corresponds to *C*). This is in contrast to *moveable do*, where solfège syllables are associated with their position on the scale (e.g., “do” corresponds to the tonic [the first, lowest

pitch] of the respective scale). It has been shown that *fixed do* training during childhood encourages the development of AP, possibly because the association between a solfège syllable and a specific pitch is continuously strengthened (Gregersen et al., 2001). Another aspect of the cultural background that has been identified to increase prevalence of AP is exposure to tone languages (Deutsch, 2002; Deutsch et al., 2009; Deutsch, Henthorn, & Dolson, 2004). The reasoning behind this observation is as follows: In tone languages, the meaning (i.e. the semantic information) of a word completely changes depending on the tone in which it is spoken (*ma* can mean both "mother" or "hemp" in Mandarin). The tone of a spoken word is determined by its AP and its pitch contour. Thus, speakers of tone languages associate the tone of a spoken word with a verbal label, similar to how AP musicians associate a pitch with its label (Deutsch, 2002). In a seminal study on this topic, Deutsch, Henthorn, and Dolson (1999) found that the tone of spoken words (in terms of its AP) is remarkably precise and consistent across days in Vietnamese and Mandarin speakers. They formulated the hypothesis that AP has evolved as a feature of speech that later generalized to musical pitch. In follow-up studies, the same research group found that a use of consistent and precise AP in spoken words was not present in English speakers (Deutsch et al., 2004), and that a better fluency in speaking a tone language was associated with higher AP ability (Deutsch et al., 2009).

Taken together, it seems clear that both genetic and environmental factors each play their roles in facilitating AP development. However, differentiating between the influences of genes and environment is a great challenge as they most certainly do not act independently but heavily interact with each other. For example, it is conceivable that there is a genetic basis for early engagement with musical stimuli which might cause an early age of onset of musical training. Furthermore, longitudinal studies that might illuminate the processes behind the development of AP are still lacking.

1.3 Two-Component Model of AP

In comparison with the scientific interest in its origins, the cognitive architecture of AP has received considerably less attention from the scientific community, although a handful of theories have been proposed (Levitin, 1994; Levitin & Rogers, 2005; McLachlan & Wilson, 2010; Ross, Gore, & Marks, 2005; Zatorre, 2003). Cognitive theories of AP are crucial to understanding what perceptual and cognitive functions might contribute to the phenomenon, which in turn helps design appropriate experiments and interpret behavioral and neural findings from these experiments. The most influential cognitive theory of AP is the two-component model (Levitin & Rogers, 2005), originally proposed by Daniel J. Levitin 25 years ago (Levitin,

1994). This model postulates that AP consists of two separate abilities – pitch memory and pitch labeling. In the following paragraphs, these abilities are discussed in detail.

1.3.1 Pitch Memory

Levitin's (1994) first component of AP, pitch memory, was inspired by studies demonstrating the presence of a latent AP form in individuals without the explicit ability to name pitches in isolation (Deutsch, 1986; Halpern, 1989; Terhardt & Seewann, 1983; Terhardt & Ward, 1982). In an exemplary study, Terhardt and Ward (1982) played participants short excerpts of classical music in their original pitch or transposed to a different pitch (up or down 1, 4, 6, or 7 semitones). The participants (n = 20 musicians) had to judge if the excerpts were played in the original or in a different pitch. Even though 90% of the participants indicated that they do not possess AP, almost all of them identified the original excerpts above chance level, and at the same level as AP musicians. Thus, these participants had the ability to recognize the correct pitches in absolute terms, even though they might not have been able to name the specific pitches. This finding has been directly and conceptually replicated in a number of studies (Smith & Schmuckler, 2008; Van Hedger, Heald, Huang, Rutstein, & Nusbaum, 2017; Van Hedger, Heald, & Nusbaum, 2016, 2018b), for example with college students who identified the original pitches of familiar theme songs from popular television shows (Schellenberg & Trehub, 2003). In a similar vein, Levitin (1994) asked participants not only to identify but to produce (e.g., sing or hum) contemporary popular songs in their original pitches. Intriguingly, almost half of the participants produced the first pitch of the songs within two semitones of the original pitch (see also Frieler et al., 2013; Jakubowski & Müllensiefen, 2013 for replications of this study).

Based on these findings, Levitin (1994) proposed that the ability to form and maintain representations of specific pitches in long-term memory is widespread in the general population. He reasoned that stable, long-term memory representations of pitches must build the basis for the explicit form of AP present in AP musicians, because only when a pitch is first identified can it be correctly named afterwards. These long-term representations might take the form of pitch templates to which the extracted pitches of incoming sounds are matched (Levitin & Rogers, 2005). Taken together, long-term pitch memory is one of two necessary components for AP, but because it is highly prevalent in the general population, it is not the component which separates individuals with and without AP.

1.3.2 Pitch Labeling

The component which does in fact differentiate AP possessors from non-possessors is called pitch labeling (Levitin, 1994; Levitin & Rogers, 2005). Pitch labeling is the ability to attach

meaningful labels to the long-term pitch representations (Levitin, 1994). These labels are predominantly in a verbal code, e.g., *C#*, *F*, or *Do* (Siegel, 1974), but there is evidence that AP musicians are able to use different coding strategies, e.g., in the visual or kinesthetic modalities (Zatorre & Beckett, 1989). The initial association between the pitch and its label is possibly formed during musical training in early childhood (Ward & Burns, 1982). Once this association is established, the label becomes part of the long-term template for this specific pitch (Levitin & Rogers, 2005). Until recently, it was assumed that the association between pitch and label is fixed once it is fully established. However, a considerable number of studies have shown that the pitch templates of AP musicians can be modified to a non-negligible extent (Athos et al., 2007; Hedger, Heald, & Nusbaum, 2013; Van Hedger, Heald, & Nusbaum, 2015; Van Hedger, Heald, Uddin, & Nusbaum, 2018; Wilson, Martin, Martin, Rayner, & McLachlan, 2012). Athos et al. (2007) were the first to empirically confirm previously reported anecdotal evidence that the perception of pitches in AP musicians systematically shifts in the "sharp" direction with advancing age (i.e. pitches are labeled higher than they actually are). Using a clever experimental design, Hedger et al. (2013) demonstrated that pitch-label associations can also be modified within much shorter timescales. They played AP musicians a digitally modified version of a classical music piece that was gradually detuned by one semitone over 30 minutes, so the detuning was not consciously noticeable. After listening, the AP musicians showed a shift in perception in the direction the musical piece was detuned. This means that the pitch templates can be modified within minutes to hours, and, thus, it follows that pitch-label associations can be weakened or strengthened on a day-to-day basis.

Pitch labeling in AP musicians is predominantly assessed using tone-naming tests (Bermudez & Zatorre, 2009; Takeuchi & Hulse, 1993). In these tests, participants are asked to label subsequently presented tones (e.g., pure tones) as accurately (and sometimes as quickly) as possible. A few studies have additionally used pitch-production tasks to assess the "active" AP ability of the participants (e.g., Dohn, Garza-Villarreal, Ribe, Wallentin, & Vuust, 2014; Wenhart, Bethlehem, Baron-Cohen, & Altenmüller, 2019). To date, there is no standard in the scientific AP community concerning stimulus selection, test administration, or scoring strategy (Bermudez & Zatorre, 2009). Interestingly, AP musicians show considerable variability in pitch labeling as assessed by tone-naming tests. This variability depends on numerous factors that have been identified in behavioral studies: First, pitch labeling varies as a function of timbre (Miyazaki, 1989; Schlemmer, Kulke, Kuchinke, & Van Der Meer, 2005; Vanzella & Schellenberg, 2010). For example, the pitches of piano tones are more easily identified than the pitches of pure tones (Miyazaki, 1989). Second, pitch labeling depends on pitch register (i.e. in

which octave the pitch is located). Pitches of very high and very low registers are identified worse than pitches of medium register (Miyazaki, 1989). Third, labeling depends on pitch class. Pitches that correspond to white keys on the piano are identified more accurately and faster than pitches that correspond to black keys (Miyazaki, 1988, 1990; Takeuchi & Hulse, 1991). Taken together, stimulus-related features (other than pitch itself) are essential for explaining variability in pitch labeling, and, consequently, these features are often held constant or are randomly varied in tone-naming tests. Another major characteristic of pitch labeling as the second component of AP is its automaticity (Levitin & Rogers, 2005). When AP musicians encounter a sound which evokes a pitch, they automatically label this pitch. Pitch-labeling automaticity has been repeatedly demonstrated using Stroop-like tasks (Akiva-Kabiri & Henik, 2012; Hsieh & Saberi, 2008b; Itoh, Suwazono, Arao, Miyazaki, & Nakada, 2005; Schulze, Mueller, & Koelsch, 2013; Stroop, 1935). In these tasks, participants are confronted with congruent stimuli, e.g., the pitch label *D* sung in the pitch of *D*, and with incongruent stimuli, e.g., the pitch label *D* sung in the pitch of *G#*. When AP musicians are asked to report the pitch label, they are typically slower when confronted with incongruent stimuli than with congruent stimuli (e.g., Itoh et al., 2005). Article 2 of this thesis includes an audio-visual version of a Stroop-like task for AP musicians (see 3.3.3).

1.4 Aims

As seen in the previous paragraphs, behavioral and (to a lesser extent) genetic studies on AP have generated interesting insights about the prevalence, origins, and behavioral aspects of the phenomenon. In contrast, the neural bases of this ability are still a mystery. One reason for this is the relatively small number of studies conducted to date. Research into the cognitive neuroscience of AP has only begun 35 years ago with a seminal EEG study published in *Science* that identified differences between AP and RP musicians in an event-related potential (ERP) component (Klein, Coles, & Donchin, 1984). Around 10 years later, another *Science* paper initiated neuroimaging research into AP by identifying a strikingly large asymmetry of the planum temporale, a secondary auditory cortical region, in AP musicians compared to RP musicians (Schlaug, Jäncke, Huang, & Steinmetz, 1995). The knowledge is further limited by the fact that many previously conducted studies on the neural bases of AP have had small sample sizes and consequently low statistical power to detect any effects that are not very large in size. The findings of these studies have a low likelihood of being robust because low power not only decreases the probability of detecting a true effect but also diminishes the probability that a detected effect represents a true effect (Button et al., 2013).

The primary aim of this thesis was to investigate the neural underpinnings of pitch processing in a large sample of AP and RP musicians. Because the ability to label pitches differentiates AP possessors from non-possessors, we focused specifically on this second component of AP. In two separate experiments, we compared AP musicians and RP musicians using EEG and fMRI during a pitch processing task with a listening and a labeling condition. The three articles included in the empirical part of this thesis are based on the data collected in these experiments.

Article 1 of this thesis directly assesses the robustness of a previous finding on pitch labeling in AP musicians. It describes an attempted replication of a study originally conducted by Itoh et al. (2005). Using EEG, the authors of the original study investigated the electrophysiological correlates of AP and RP. They found an AP-specific ERP component elicited during simple tone listening and several ERP components related to RP during labeling. In our study, we aimed to replicate the main findings of Itoh et al.'s study using closely matched materials and methods, but with more than double the sample size.

In article 2, we reanalyzed the same EEG data using a state-of-the-art "brain decoding" framework to investigate the neural representations underlying listening and labeling. We aimed to test a prediction that follows from the automaticity of pitch labeling, namely that listening and labeling are more similar in AP musicians than in RP musicians in terms of their neural representations because AP musicians (but not RP musicians) automatically label pitches even when not instructed to.

Article 3 describes the findings from an fMRI experiment during which we presented pure tones and noise segments to the participants. The participants were asked to either listen or label the pitches of the tones. Here, we aimed to uncover which and how brain regions contribute to pitch labeling in AP musicians. We comprehensively analyzed the fMRI data using the general linear model (GLM), multivariate pattern analysis (MVPA), and a region of interest (ROI) analysis.

Chapter 2 Empirical Work – Article 1

A reevaluation of the electrophysiological correlates of absolute pitch and relative pitch: no evidence for an absolute pitch-specific negativity

Simon Leipold, Chantal Oderbolz, Marielle Greber, & Lutz Jäncke

2.1 Abstract

Musicians with absolute pitch effortlessly identify the pitch of a sound without an external reference. Previous neuroscientific studies on absolute pitch have typically had small samples sizes and low statistical power, making them susceptible for false positive findings. In a seminal study, Itoh et al. (2005) reported the elicitation of an absolute pitch-specific event-related potential component during tone listening — the *AP negativity*. Additionally, they identified several components as correlates of relative pitch, the ability to identify relations between pitches. Here, we attempted to replicate the main findings of Itoh et al.'s study in a large sample of musicians (n = 104) using both frequentist and Bayesian inference. We were not able to replicate the presence of an AP negativity during tone listening in individuals with high levels of absolute pitch, but we partially replicated the findings concerning the correlates of relative pitch. Our results are consistent with several previous studies reporting an absence of differences between musicians with and without absolute pitch in early auditory evoked potential components. We conclude that replication studies form a crucial part in assessing extraordinary findings, even more so in small fields where a single finding can have a large impact on further research.

This article was originally published in:

Leipold, S., Oderbolz, C., Greber, M., & Jäncke, L. (2019). A reevaluation of the electrophysiological correlates of absolute pitch and relative pitch: no evidence for an absolute pitch-specific negativity. *International Journal of Psychophysiology*, *137*, 21-31. https://doi.org/10.1016/j.ijpsycho.2018.12.016

2.2 Introduction

A core pillar of scientific progress is the validation of the veracity of findings by means of replication studies. If an effect is true, it should be reliably obtained in an independent, adequately powered study using similar procedures (Simons, 2014). In recent years, the replicability of published findings has been repeatedly called into question. Meta-research studies have demonstrated, both theoretically and empirically, low replicability of findings across a diverse range of research fields (Baker, 2016; Begley & Ellis, 2012; Camerer et al., 2016, 2018; Ioannidis, 2005; Open Science Collaboration, 2015). Several reasons contributing to the low replicability of findings have been identified (Munafò et al., 2017). These include, for example, publication bias against negative results (Nissen, Magidson, Gross, & Bergstrom, 2016; Rosenthal, 1979), questionable research practices (Kerr, 1998; Neuroskeptic, 2012; Simmons, Nelson, & Simonsohn, 2011), misunderstandings of *p* values (Halsey, Curran-Everett, Vowler, & Drummond, 2015), and maybe most importantly, low statistical power (Button et al., 2013).

Not only does low statistical power result in a decreased probability of finding an effect should a true effect exist, it also reduces the probability of a significant effect being a true effect (Button et al., 2013). Cognitive neuroscience represents a research field wherein the problem of low statistical power is especially evident (Nord, Valton, Wood, & Roiser, 2017; Szucs & Ioannidis, 2017). The use of non-invasive techniques such as functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) on human subjects is both time-consuming and resource-intensive; thus, cognitive neuroscience studies often have small sample sizes which are only adequate to detect very large effects (Poldrack et al., 2017). In addition, the problem of small samples is aggravated in studies investigating rare populations with unique characteristics because the subject pool of these studies is inherently limited.

A prime example for studies on rare populations using small samples sizes are neuroscientific studies on individuals with absolute pitch (AP) — the ability to effortlessly name the pitch of a sound without an external reference sound (Deutsch, 2013). In the last 25 years, these studies have provided important insights into the cognitive, neuroanatomical, and neurophysiological foundations of this ability (Keenan, Thangaraj, Halpern, & Schlaug, 2001; Schlaug et al., 1995; Zatorre, Perry, Beckett, Westbury, & Evans, 1998). However, many of the findings have not yet been replicated. In 2005, a seminal EEG study on AP was published in *Cerebral Cortex* (Itoh et al., 2005). The authors evaluated four groups of subjects with varying levels of AP (High-AP musicians, Mid-AP musicians, Low-AP musicians, and untrained individuals without AP). Each of the groups consisted of 11 subjects. The main finding of Itoh et al.'s study

was the unique elicitation of a left posterior temporal negativity in High-AP musicians in response to pure tones. This component of the event-related potential (ERP) occurring 150 ms after stimulus-onset was termed *AP negativity* because it presumably reflected the automatic retrieval of the association of a pitch and its label (e.g., C#) in High-AP musicians. Further findings of Itoh et al.'s study concerned the ERP correlates of relative pitch (RP), the ability to identify the relation between successive pitches either by making higher-lower judgments or by determining the exact musical interval between the pitches (McDermott & Oxenham, 2008). The authors found three ERP components (P3b, parietal positive slow wave, frontal negative slow wave) occurring later than 300 ms after stimulus-onset that were not elicited in High-AP musicians and gradually increased with lower levels of AP. Thus, these components presumably reflected the cortical processing related to RP.

In small research fields such as those exploring AP, single studies with extraordinary findings can have a large impact on subsequently conducted studies. Itoh et al.'s study, a prime example of such a case, was the first to find an electrophysiological marker of AP occurring as early as 150 ms after stimulus-onset during simple tone listening. Subsequent studies have cited this aspect of the study as part of the motivation for their use of EEG to investigate the timing of the neurophysiological responses in AP (e.g., Elmer, Sollberger, Meyer, & Jäncke, 2013; Rogenmoser, Elmer, & Jäncke, 2015). Furthermore, many studies cite that the AP negativity was (exclusively) found at an electrode over the left temporal cortex as evidence for the importance of the left-sided planum temporale in AP (e.g., Loui, Zamm, & Schlaug, 2012; Oechslin, Meyer, & Jäncke, 2010; Schulze, Gaab, & Schlaug, 2009). Surprisingly, since the publication of Itoh et al.'s study more than 10 years ago, no study has been able to replicate the finding of an AP negativity in musicians with high levels of AP. This is in stark contrast to the findings concerning the ERP correlates of RP, which are consistent with numerous studies finding an absent or reduced P3b component in AP musicians (Hantz, Crummer, Wayman, Walton, & Frisina, 1992; Klein et al., 1984; Wayman, Frisina, Walton, Hantz, & Crummer, 1992).

In several recent high-profile publications, the need for replications and its scientific value has been strongly emphasized (Munafò et al., 2017; Poldrack et al., 2017; Zwaan, Etz, Lucas, & Donnellan, 2018). There have been first attempts to replicate findings in cognitive neuroscience (e.g., Boekel et al., 2015; Nieuwland et al., 2018). In light of these developments, we attempted to replicate the main findings of the study by Itoh et al. using closely matched materials, procedures, and analyses. We recruited a sample of 104 musicians — more than double the sample size of the original study. These subjects performed a pitch-naming task during EEG

acquisition. We evaluated the AP negativity as well as the ERP components related to RP, namely the P3b, the parietal positive slow wave (ppSW), and the frontal negative slow wave (fnSW). To quantify the success of the replication, we used Bayesian inference along with the frequentist inference employed in the original study.

2.3 Materials and Methods

2.3.1 Subjects

We examined 104 musicians with varying degrees of AP that were assigned to three groups (High-AP musicians, Mid-AP musicians, Low-AP musicians) according to their tone-naming proficiency (see Figure 1). The group assignment was based on two cutoffs in tone-naming scores (37.5%, 82.5%) that were chosen to optimize the balance between, on the one hand, a close matching of the group-specific tone-naming scores in the original study and, on the other hand, a high enough number of subjects in each group. Note that the cutoff scores chosen are specific to the sample investigated in this study and should not be taken as non-arbitrary benchmarks for future studies. In the original study, the criterion for group assignment was not reported. Unlike in the original study, we did not include a group without musical training for two reasons: It would have been difficult for them to understand the instructions of the EEG task, and their results would have been hard to interpret because of the confounding factor of musical training.

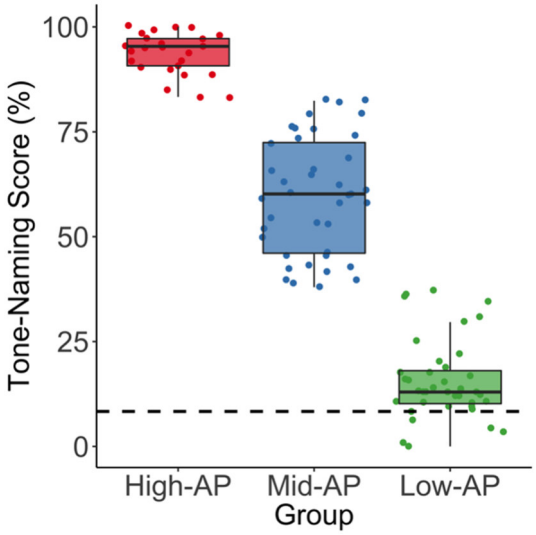


Figure 1. Distribution of tone-naming scores across groups.

The subjects (n = 104) were assigned to three groups — High-AP musicians, Mid-AP musicians, and Low-AP musicians — according to their tone-naming proficiency. The assignment was based on two cutoffs in tone-naming scores (37.5%, 82.5%). These values were chosen to optimize the balance between a close matching of the group-specific tone-naming scores in the original study and a sufficient number of subjects in each group.

All subjects were either professional musicians, music students, or highly trained amateurs between 18 and 44 years of age, who were recruited in the context of a larger project investigating AP. None of the subjects reported any neurological, audiological, or severe psychiatric disorders. Pure tone audiometry (ST20, MAICO Diagnostics, Berlin, Germany) confirmed the absence of hearing loss in all subjects. The demographical data (sex, age, handedness) and part of the behavioral data (tone-naming proficiency and musical experience) were collected using an online survey tool (http://www.limesurvey.org/). Self-reported handedness was verified using a German translation of the Annett questionnaire (Annett, 1970). Subjects provided written informed consent and were paid for their participation. The study was approved by the local ethics committee (http://www.kek.zh.ch/) and conducted according to the principles defined in the Declaration of Helsinki.

2.3.2 Tone-Naming Test

To assess their tone-naming ability, all subjects completed a tone-naming test (Oechslin et al., 2010). During the test, the subjects had to name both the chroma and the octave (e.g. E4) of 108 pure tones which were presented in a pseudorandomized order. The tones had a duration of 500 ms and were masked with 2000 ms Brownian noise presented immediately before and after each tone. The test included all tones from C3 to B5 (twelve-tone equal temperament tuning, A4 = 440 Hz). Each of the tones was presented three times. The tone-naming score was calculated using the percentage of correct chroma identifications without considering octave errors (Deutsch, 2013). Therefore, the chance-level was at 8.3%.

2.3.3 Experimental Procedure

During EEG data acquisition, the subjects performed a slightly modified version of the pitch-naming task employed in the original study. Itoh et al. also included a Stroop-like ERP experiment in their study which we chose not to replicate because in the original study, it mainly served as an attempted internal replication of the identified ERP correlates of AP and RP. Thus, in this study, we do not discuss the findings of this Stroop-like experiment.

All stimuli and the stimulus presentation scripts are available online (https://osf.io/93gdm/). We used identical auditory stimuli to the ones used in the original study. The stimuli consisted of three pure tones with different frequencies (262 Hz, 294 Hz, and 330 Hz), corresponding to C4, D4, and E4 in twelve-tone equal temperament tuning. Note that in Itoh et al.'s publication, the tones were labeled incorrectly (C3, D3, and E3). All tones had the same temporal envelope, characterized by a duration of 350 ms, a 10 ms linear fade-in, and a 50 ms linear fade-out. The stimuli were created using Audacity (version 2.1.2, http://www.audacityteam.org/) and

presented at a sound pressure level of 75 dB via on-ear headphones (HD 25-1, Sennheiser, Wedemark, Germany) using Presentation software (version 18.1, www.neurobs.com).

The pitch-naming task consisted of a Listening and a Labeling condition (the Naming condition of the original study). Both experimental conditions encompassed 180 trials, and each of the three pure tones was presented 60 times. The trials were presented in a randomized order. Within a single trial, first, a pure tone was presented, followed by a silent inter-stimulus interval (jittered duration = 900 ms – 1100 ms) after which an auditory cue (pink noise, duration = 10 ms, linear fade-in = 2 ms, linear fade-out = 2 ms) was presented. This cue indicated to the subjects that they should respond by a key press. The trial continued in silence until a response was given. After the response, there was an inter-trial interval (duration = 1000 ms), before the next pure tone was presented. The experimental conditions only differed in terms of the instructions given to the subjects: In the Listening condition, the subjects listened to the tones and pressed a neutrally marked key in response to the auditory cue, irrespective of the chroma of the pure tone that was presented before. In the Labeling condition, the subjects labeled the pure tones by pressing one of three corresponding keys that were marked with the tone names (C, D, and E), also in response to the auditory cue. In both conditions, the subjects were instructed to respond as quickly and as accurately as possible. Importantly, no responses were given directly after the tone in order to avoid a contamination of the ERPs by motor artifacts. Before each condition, we included six practice trials to familiarize the subjects with the task. During both experimental conditions, a screen in front of the subjects showed a black fixation cross on a gray background. The whole task lasted around 20 minutes.

Our version of the pitch-naming task differed from Itoh et al.'s version in three ways: First, we doubled the number of trials per condition (180 instead of 90) to increase the signal to noise ratio and therefore the statistical power. Second, in order to accurately record the subjects' responses and to avoid articulation-related artifacts, we chose to refrain from using verbal responses. In the original study, the subjects responded verbally. Third, we fixed the order of the conditions across all subjects whereas in the original study, the order was balanced across subjects. The rationale for this decision is as follows: Group differences in ERP amplitudes during Listening are interpreted in relation to the AP-specific automaticity of the retrieval of the pitch-label association that occurs even without external instruction. Consequently, in order to avoid the contamination of the Listening condition by possible spill-overs from the instruction of a preceding Labeling condition, we decided to always present the Listening condition before the Labeling condition.

2.3.4 EEG Data Acquisition and Preprocessing

The EEG data was acquired continuously using an electrode cap (Easycap, Herrsching, Germany) with 32 Ag / AgCl electrodes placed according to an extended 10/20 system (Fp1, Fp2, F7, F3, Fz, F4, F8, FT7, FC3, FCz, FC4, FT8, T7, C3, Cz, C4, T8, TP9, TP7, CP3, CPz, CP4, TP8, TP10, P7, P3, Pz, P4, P8, O1, Oz, O2) and a BrainAmp amplifier (Brainproducts, Munich, Germany). The reference electrode was placed on the tip of the nose. We used a sampling rate of 1000 Hz and an online bandpass filter between 0.1 Hz and 100 Hz. Electrode impedance was kept below 10 kΩ throughout the acquisition using electrically conductive gel.

The preprocessing of the EEG data was performed using BrainVision Analyzer (Version 2.1, https://www.brainproducts.com/). We bandpass-filtered the data from 0.5 Hz to 20 Hz (48 dB/octave) and applied a notch filter of 50 Hz. Artifacts caused by eye blinks and saccades were corrected using independent component analysis (Jung et al., 2000) with the remaining artifacts being removed using an automatic raw data inspection (removal criteria: amplitude gradient > 50 μV/ms, amplitude difference > 100 μV, amplitude minimum/maximum > -100 μV / 100 μV). Afterwards, we segmented the continuous data into epochs of 900 ms (-100 ms to 800 ms relative to pure-tone onset). Finally, we baseline-corrected these epochs using the interval from -100 ms to the onsets of the pure tones.

The baseline-corrected epochs were averaged per subject and experimental condition to compute ERPs. From these ERPs, we extracted the mean amplitudes of the following time intervals, originally specified in Itoh et al.'s study: P3b = 300 – 450 ms, ppSW = 450 – 550 ms, fnSW = 550 ms – 800 ms, AP negativity = 145 – 155 ms. Note that for the fnSW, Itoh et al. used an interval from 550 ms to 900 ms (the end of their epochs). In our case, however, due to the jittered inter-stimulus interval, some of the auditory cues already appeared 900 ms after the onset of the pure tones and thus, to avoid motor artifacts from key presses in anticipation of the auditory cue, we reduced the designated interval by 100 ms. The extracted mean amplitudes were then subjected to statistical analysis. We only analyzed 20 of the 32 electrodes to match the electrode locations used in the original study. The analyzed electrodes were as follows: Fp1, Fp2, F7, F3, Fz, F4, F8, T7, C3, Cz, C4, T8, P7, P3, Pz, P4, P8, O1, Oz, O2. Note that in the original study, the electrode locations T7, T8, P7, and P8 are called T3, T4, T5, and T6 respectively, in line with the classical 10-20 system terminology. Thus, in the following, we refer to these electrode locations as T7/T3, T8/T4, P7/T5 and P8/T6.

2.3.5 Statistical Analysis

The statistical analysis of the EEG data and the behavioral data was performed in R (version 3.3.2, http://www.r-project.org/). In addition to the frequentist inference employed in the original study, we used Bayesian inference, more specifically Bayes factors, to quantify the evidence for the alternative hypothesis relative to the null hypothesis and vice versa (Kass & Raftery, 1995). In contrast to frequentist inference, which can only be used to reject the null hypothesis, Bayesian inference allows statements concerning the evidence in support of the null hypothesis (Dienes, 2011). As such, Bayesian inference is ideally suited for replication studies to avoid the non-interpretability of a non-significant effect obtained using frequentist inference (Anderson & Maxwell, 2016). Consequently, in addition to *p* values, we report Bayes factors favoring either the null hypothesis (BF_{01}) or the alternative hypothesis (BF_{10}). Bayes factors are interpreted in a straightforward way: A BF_{10} of 3 indicates that the observed data is three times more likely under the alternative hypothesis than under the null hypothesis. In this study, a Bayes factor between 1 and 3 is considered as anecdotal evidence, a Bayes factor between 3 and 10 as moderate evidence, and a Bayes factor between 10 and 30 as strong evidence for one hypothesis relative to the other hypothesis (Boekel et al., 2015; Jeffreys, 1961). To calculate Bayesian t-tests (Rouder, Speckman, Sun, Morey, & Iverson, 2009) and Bayesian ANOVAs (Rouder, Morey, Speckman, & Province, 2012), we used the R package *BayesFactor* (version 0.9.12-2, https://CRAN.R-project.org/package=BayesFactor). We used the default priors as implemented in the *BayesFactor* package (scale value $r = 0.707$). The use of default priors is advantageous because these priors do not depend on effect size estimates drawn from previous studies which are often known to be inflated, especially in studies with small sample sizes (Ioannidis, 2008). Nonetheless, we checked a range of different scale values, which did not change the conclusions drawn from the resulting Bayes factors. Hence, only Bayes factors based on the default scale value are reported. To calculate frequentist ANOVAs, we used the R package *ez* (version 4.4.0, https://CRAN.R-project.org/package=ez). In case of non-sphericity, Greenhouse-Geisser corrected degrees of freedom and *p* values are reported. The significance level was set to $\alpha = 0.05$ for all analyses unless otherwise stated. Effect sizes within an ANOVA are given as generalized eta-squared (η^2_G) and effect sizes for t-tests are given as Cohen's d (*d*).

2.3.6 Behavioral Data Analysis

The behavioral and demographical subject characteristics were compared between the three groups using a one-way ANOVA per characteristic. The behavioral measures acquired during the EEG experiment (response accuracy and response time) were analyzed separately. For the response time, we calculated a two-way mixed-design ANOVA with a within-subject factor

Condition and a between-subject factor Group. Trials with response times shorter than 200 ms or longer than 1000 ms were excluded from the analysis. For the response accuracy in the Labeling condition, we calculated a one-way ANOVA with a between-subject factor Group and subsequently performed pairwise comparisons (Bonferroni adjusted $\alpha = 0.017$). In the Listening condition, there was no response choice involved and thus, no response accuracy was calculated.

2.3.7 EEG Data Analysis

For each of the ERP components related to RP processing — P3b, ppSW, and fnSW — we first calculated a two-way mixed-design ANOVA with a within-subject factor Condition (Listening vs. Labeling) and a between-subject factor Group (High-AP vs. Mid-AP vs. Low-AP) at the electrode locations identified in the original study as showing a statistically significant Group x Condition interaction. Consequently, we analyzed the amplitudes of both the P3b and the ppSW at electrode Pz. As we did not collect data for electrode Fpz, we calculated and analyzed the mean amplitudes of electrodes Fp1 and Fp2 for the fnSW. In case of a statistically significant Group x Condition interaction, for each group we calculated the difference in amplitudes between the conditions (Labeling minus Listening). These differences were then compared pairwise between the groups using two sample t-tests (Bonferroni adjusted $\alpha = 0.017$). Note that in the original study, differential condition effects between the groups were inferred based on a statistically significant condition effect in one group and a simultaneous non-significant condition effect in another group. Although somewhat common in earlier neuroscientific research, this procedure is erroneous in that a comparison of condition effects requires a statistical test on their difference (Gelman & Stern, 2006; Nieuwenhuis, Forstmann, & Wagenmakers, 2011). Nevertheless, to ensure comparability with the original study, we additionally implemented this procedure in order to check whether a component was elicited in one group, but not in another. This was done by comparing the amplitudes of the two conditions within each group separately using a paired t-test (Bonferroni adjusted $\alpha = 0.017$).

To explore potential effects at unexpected electrodes and to ensure comparability with the original study, for each RP-related component, we additionally calculated a three-way mixed-design ANOVA with the two within-subject factors Condition and Electrode (all 20 electrodes) and a between-subject factor Group. We did not calculate Bayes factors for the three-way ANOVAs because they are not well-suited for designs of high complexity (Rouder, Morey, Verhagen, Swagman, & Wagenmakers, 2017). In case of a three-way interaction, a two-way ANOVA with the factors Condition and Group was performed at each of the 20 electrodes (Bonferroni adjusted $\alpha = 0.0025$). We used the conservative Bonferroni correction because in

an N-way ANOVA with a large number of factor levels, the family-wise error (FWE) rate is already inflated at the level of the ANOVA itself (Cramer et al., 2016; Luck & Gaspelin, 2017). Therefore, in follow-up ANOVAs, it is crucial to adequately correct for the inflated FWE rate. For electrodes showing a statistically significant Group x Condition interaction, we again calculated condition differences and compared these differences pairwise between the groups using two sample t-tests (Bonferroni adjusted $\alpha = 0.017$). Finally, we again checked whether a component was elicited in one group, but not another. It is important to note that it is generally not advisable to compute N-way ANOVAs with a large number of factor levels as it was done in the original study. As described above, on the one hand, in these types of ANOVAs the FWE rate and thus the probability for a false-positive finding increases. On the other hand, an adequate adjustment of the significance level to correct for the inflated FWE rate decreases the statistical power of the study (Luck & Gaspelin, 2017).

The ERP component related to AP processing (AP negativity) was only analyzed in the Listening condition at electrode P7/T5 using a one-way ANOVA with a between-subject factor Group. In the original study, group differences were again erroneously inferred based on a significant effect in one group and a simultaneous non-significant effect in the other group (Gelman & Stern, 2006; Nieuwenhuis et al., 2011). To ensure comparability with the original study, we additionally checked in which of the groups the component was elicited above baseline, using a t-test against zero in each group (Bonferroni adjusted $\alpha = 0.017$).

2.4 Results

2.4.1 Behavior

The three groups did not differ with regard to age ($F(2,101) = 1.46$, $p = 0.24$, $BF_{01} = 3.28$, $\eta^2_G = 0.03$) and years of musical training ($F(2,101) = 1.47$, $p = 0.24$, $BF_{01} = 3.31$, $\eta^2_G = 0.03$). As expected, the three groups substantially differed in their tone-naming proficiency ($F(2,101) = 422.12$, $p < 10\text{-}49$, $BF_{10} > 1046$, $\eta^2_G = 0.89$). High-AP musicians had a higher tone-naming score than Mid-AP musicians ($t(63) = 11.65$, $p < 0.001$, $BF_{10} > 1014$, $d = 2.97$) and Mid-AP musicians had a higher score than Low-AP musicians ($t(77) = 16.36$, $p < 0.001$, $BF_{10} > 1023$, $d = 3.68$). Note that both in the original study and in this study, the subjects of the Low-AP group performed only slightly above chance-level. Therefore, these subjects represent what previous studies have termed Non-AP or RP musicians. To be consistent with the terminology of the original study, we are also using the term Low-AP musicians. Descriptive statistics of the subject characteristics are given in Table 1.

Table 1. Subject characteristics.

Continuous measures given as mean ± standard deviation.

	High-AP musicians	Mid-AP musicians	Low-AP musicians
Number of subjects	25	40	39
Sex (female / male)	12 / 13	19 / 21	20 / 19
Handedness (right / left / both)	23 / 1 / 1	34 / 4 / 2	35 / 3 / 1
Age (years)	25.80 ± 4.74	27.02 ± 5.72	25.10 ± 4.45
Tone-naming score (%)	93.59 ± 5.02	59.70 ± 13.96	15.69 ± 9.45
Training (years)	20.40 ± 4.54	20.62 ± 6.80	18.62 ± 4.69

Abbreviations: AP = absolute pitch

The two-way ANOVA of the response time revealed no main effect of Group ($F(2,101) = 1.69$, $p = 0.19$, $BF_{01} = 1.66$, $\eta^2_G = 0.03$), no main effect of Condition ($F(1,101) = 0.001$, $p = 0.98$, $BF_{01} = 6.77$, $\eta^2_G < 0.001$), and no Group x Condition interaction ($F(2,101) = 0.46$, $p = 0.63$, $BF_{01} = 7.94$, $\eta^2_G = 0.001$). Because the subjects were instructed to press a key in response to a delayed cue and not in response to the tones itself, these response times should only be interpreted as markers of attentional or motivational processes and not with regard to the processing of the tones itself. Accordingly, there were no systematic differences between the groups or the conditions concerning the subjects' attention or motivation during the experiment. The one-way ANOVA of the response accuracy in the Labeling condition revealed a main effect of Group ($F(2,101) = 4.62$, $p = 0.01$, $BF_{10} = 3.90$, $\eta^2_G = 0.08$), in which Mid-AP musicians showed a higher response accuracy than Low-AP musicians ($t(77) = 2.78$, $p = 0.007$, $BF_{10} = 6.16$, $d = 0.63$). However, High-AP musicians did not show a higher response accuracy than Mid-AP musicians ($t(63) = -0.99$, $p = 0.32$, $BF_{01} = 2.55$, $d = 0.25$) and Low-AP musicians ($t(62) = 1.63$, $p = 0.11$, $BF_{01} = 1.27$, $d = 0.42$). Consistent with the response accuracies reported in the original study, all groups performed on a high level: mean (± standard deviation) of High-AP = 99.56% (± 0.89), Mid-AP = 99.74% (± 0.58), and Low-AP = 98.97% (± 1.63).

2.4.2 Electrophysiological Correlates of Relative Pitch

The group-averaged ERPs per condition at electrodes Fz, Cz, and Pz are shown in Figure 2. Condition differences are clearly visible in later ERP components (starting from 300 ms after pure tone onset). Group-averaged difference waveforms (Labeling minus Listening) and difference topographies for the P3b and the ppSW components are shown in Figure 3. In all groups, a P3b and a ppSW are clearly discernible from the difference waveforms. Both components were maximally elicited at parietal electrodes. Group-averaged difference

waveforms for the fnSW component are shown in Figure 4A. The ERPs and topographies were visualized using functions from the R package *eegUtils* (Craddock, 2018).

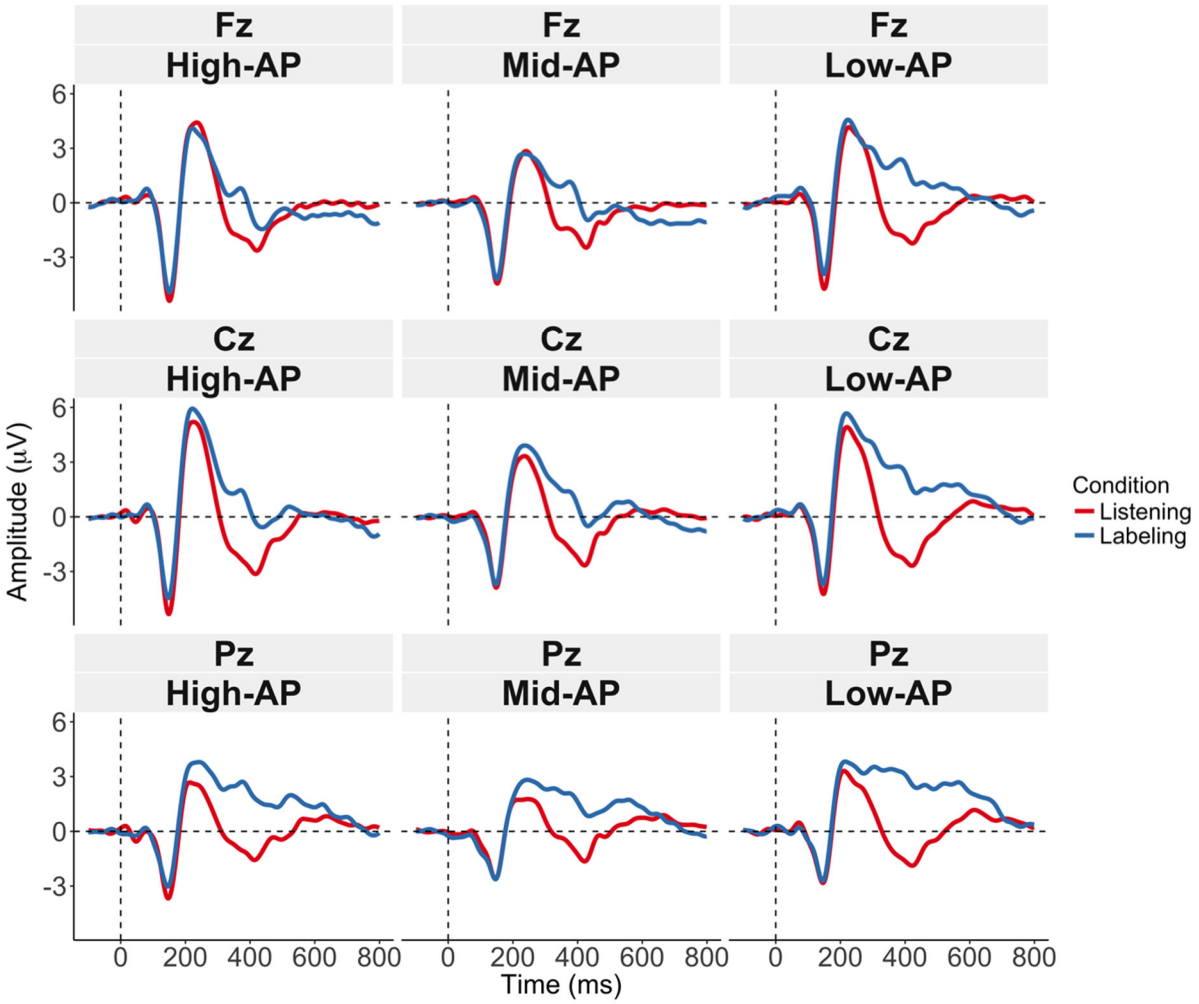


Figure 2. Event-related potential waveforms.

Group-averaged ERP waveforms per experimental condition at electrodes Fz (upper panels), Cz (middle panels), and Pz (lower panels). The ERPs show the typical characteristics of an auditory evoked potential, including an N1-P2 complex. Condition differences are clearly visible in components elicited later than 300 ms after pure-tone onset.

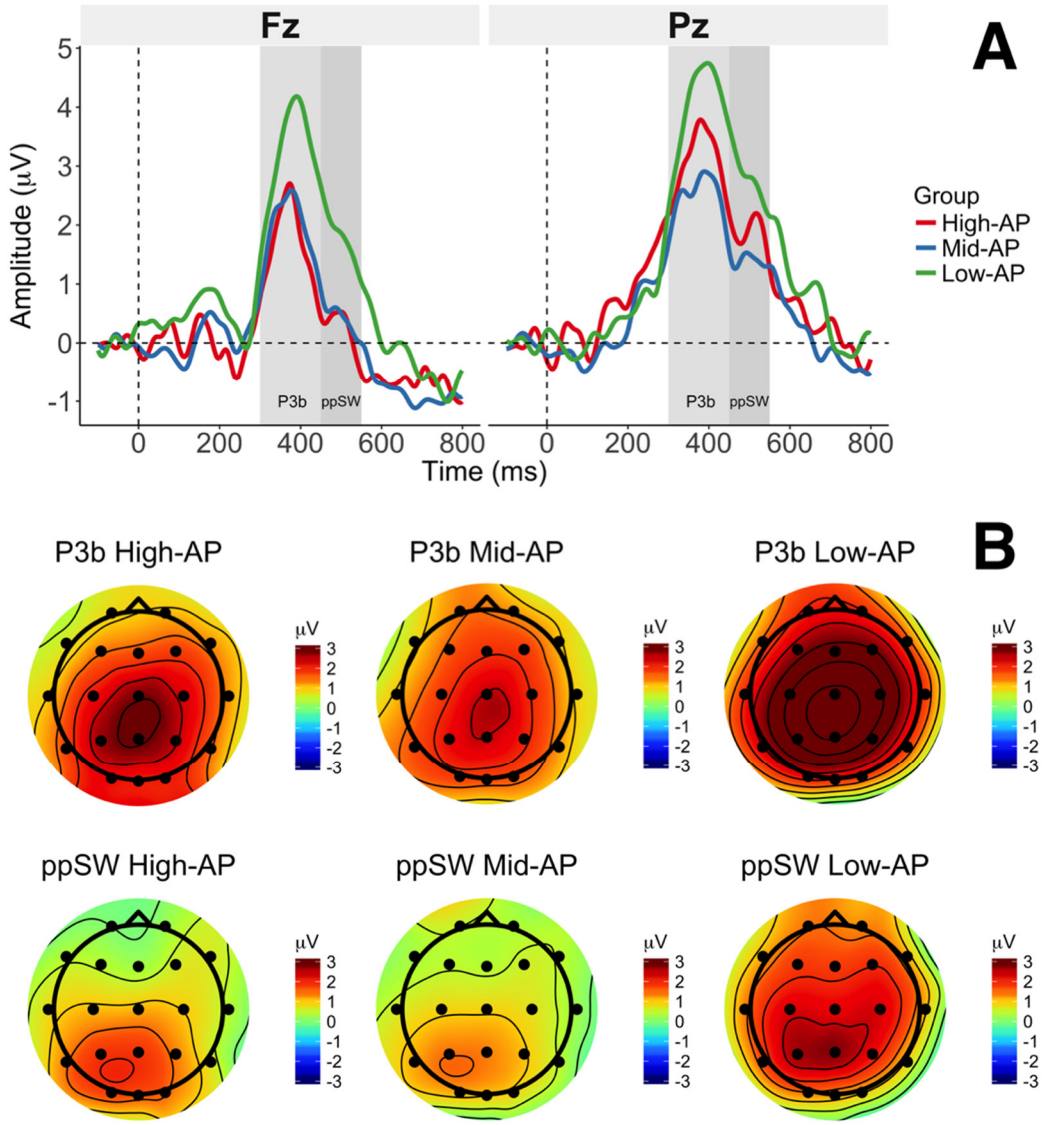


Figure 3. Replicated ERP correlates of relative pitch.

Difference waveforms and topographies (Labeling minus Listening) showing the P3b and parietal positive slow wave (ppSW) components. (A) The group-specific difference waveforms (Labeling minus Listening) showed clearly discernible P3b (light gray-shaded) and ppSW (dark gray-shaded) components. At electrode Fz, Low-AP musicians showed a larger P3b than Mid-AP musicians and High-AP musicians. There were no differences between Mid-AP and High-AP musicians. The ppSW was not analyzed at Fz because the two-way interaction (Group x Condition) at this electrode did not reach statistical significance. At electrode Pz, Low-AP musicians showed a larger P3b than Mid-AP musicians, but we found no P3b differences between Low-AP and High-AP musicians and between Mid-AP and High-AP musicians. Furthermore, Low-AP musicians showed a larger ppSW than Mid-AP musicians, but there were no ppSW differences between Low-AP and High-AP musicians and between Mid-AP and High-AP musicians. (B) The group-specific difference topographies (Labeling minus Listening) showed a maximal elicitation of the P3b and the ppSW over parietal electrodes in all groups. Low-AP musicians showed a more extensive elicitation of both components.

P3b

The two-way ANOVA of the P3b amplitude at electrode Pz revealed a Group x Condition interaction ($F(2,101) = 3.86$, $p = 0.02$, $BF_{10} = 2.31$, $\eta^2_G = 0.02$). The pairwise comparisons of condition differences in the P3b revealed that Low-AP musicians showed a larger difference than Mid-AP musicians ($t(77) = 2.96$, $p = 0.004$, $BF_{10} = 9.38$, $d = 0.67$). In contrast, there were no differences between Low-AP and High-AP musicians ($t(62) = 1.32$, $p = 0.19$, $BF_{01} = 1.84$, $d = 0.34$) and between Mid-AP and High-AP musicians ($t(63) = -1.00$, $p = 0.32$, $BF_{01} = 2.53$, $d = 0.25$). At electrode Pz, the P3b component was elicited in Low-AP musicians ($t(38) = 9.46$, $p < 0.001$, $BF_{10} > 108$, $d = 1.51$), in Mid-AP musicians ($t(39) = 8.33$, $p < 0.001$, $BF_{10} > 107$, $d = 1.32$), and in High-AP musicians ($t(24) = 5.26$, $p < 0.001$, $BF_{10} = > 103$, $d = 1.05$).

The three-way ANOVA of the P3b amplitude using all electrodes revealed a Group x Condition x Electrode interaction ($F(6.41,323.46) = 2.72$, $p = 0.01$, $\eta^2_G = 0.002$). The follow-up two-way ANOVAs revealed a Group x Condition interaction at electrode Fz ($F(2,101) = 6.67$, $p = 0.002$, $BF_{10} = 18.81$, $\eta^2_G = 0.01$). The pairwise comparison of condition differences revealed that Low-AP musicians showed a higher difference than Mid-AP musicians ($t(77) = 3.13$, $p = 0.002$, $BF_{10} = 14.32$, $d = 0.71$) and Low-AP musicians also showed a higher difference than High-AP musicians ($t(62) = 2.98$, $p = 0.004$, $BF_{10} = 9.51$, $d = 0.76$). However, there were no differences between Mid-AP musicians and High-AP musicians ($t(63) = 0.52$, $p = 0.60$, $BF_{01} = 3.43$, $d = 0.13$). At electrode Fz, again, the component was elicited in Low-AP musicians ($t(38) = -10.21$, $p < 0.001$, $BF_{10} > 109$, $d = 1.64$), in Mid-AP musicians ($t(39) = -7.32$, $p < 0.001$, $BF_{10} > 106$, $d = 1.16$), and in High-AP musicians ($t(24) = -4.08$, $p < 0.001$, $BF_{10} = 73.29$, $d = 0.82$).

Parietal Positive Slow Wave (ppSW)

The two-way ANOVA of the ppSW amplitude at electrode Pz revealed a Group x Condition interaction ($F(2,101) = 3.30$, $p = 0.04$, $BF_{10} = 1.41$, $\eta^2_G = 0.02$). According to the pairwise comparisons of condition differences, Low-AP musicians showed a larger condition difference in ppSW amplitude than Mid-AP musicians ($t(77) = 2.77$, $p = 0.007$, $BF_{10} = 6.04$, $d = 0.62$). In contrast there were no differences between Low-AP and High-AP musicians ($t(62) = 1.32$, $p = 0.19$, $BF_{01} = 1.85$, $d = 0.34$) and between Mid-AP and High-AP musicians ($t(63) = -0.80$, $p = 0.43$, $BF_{01} = 2.95$, $d = 0.20$). The ppSW component was elicited in all groups: Low-AP ($t(38) = 7.02$, $p < 0.001$, $BF_{10} > 105$, $d = 1.12$), Mid-AP ($t(39) = 4.31$, $p < 0.001$, $BF_{10} = 229.35$, $d = 0.68$), and High-AP ($t(24) = 3.09$, $p = 0.005$, $BF_{10} = 8.64$, $d = 0.62$). The three-way ANOVA of the ppSW amplitude using all electrodes revealed a Group x Condition x Electrode interaction

($F(6.06,306.10) = 2.89$, $p = 0.01$, $\eta^2_G = 0.004$). However, the follow-up two-way ANOVAs did not reveal a Group x Condition at any electrode (all $p >$ Bonferroni adjusted α).

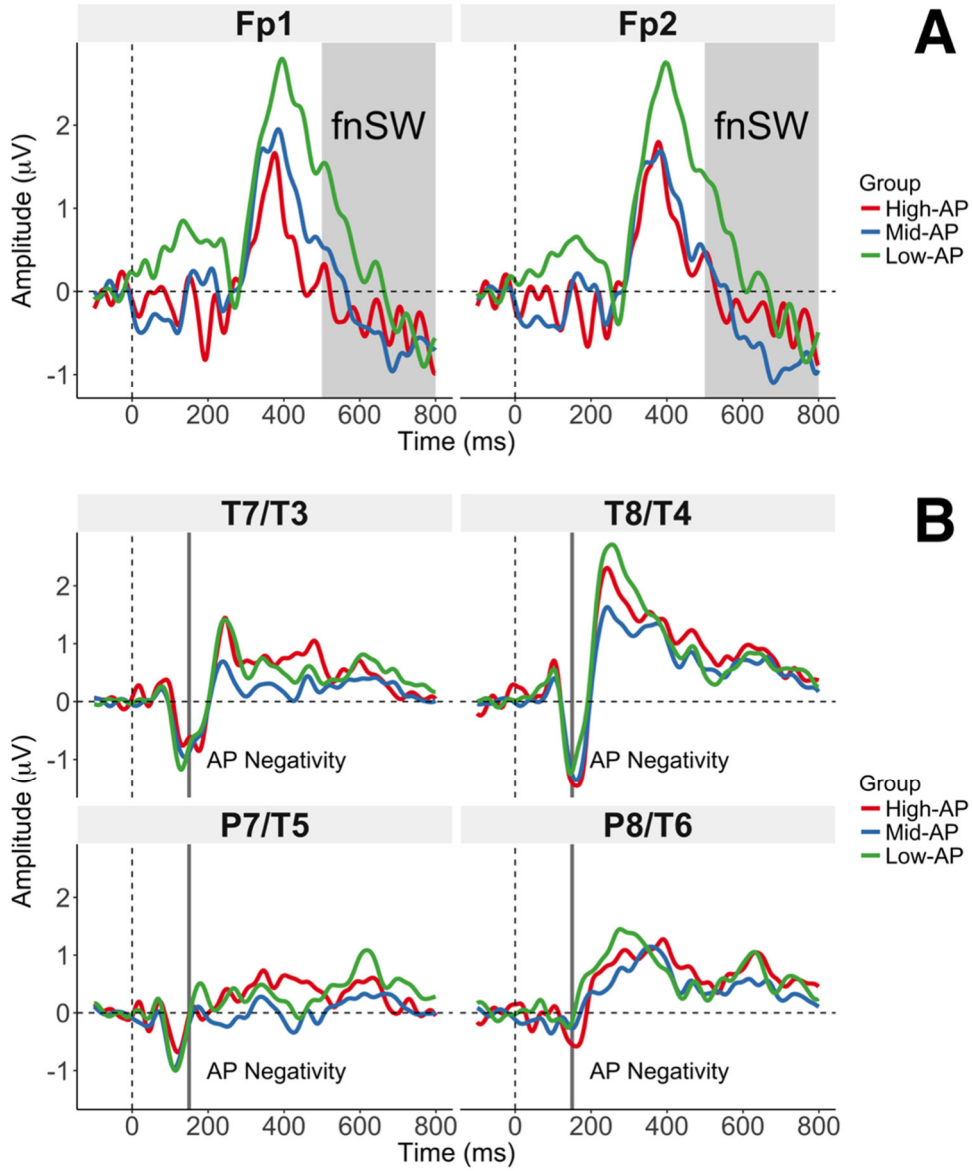


Figure 4. Non-replicated ERP correlates of relative and absolute pitch.

(A) Group-averaged difference waveforms (Labeling minus Listening) at electrodes Fp1 and Fp2 showing the frontal negative slow wave (fnSW). In contrast to the original study, we found no group differences in the fnSW (gray-shaded area). Note that we analyzed the fnSW as the mean of Fp1 and Fp2 as we did not collect data at electrode location Fpz. (B) Group-averaged waveforms in the Listening condition at medial temporal (T7/T3, T8/T4) and posterior temporal electrodes (P7/T5, P8/T6). Contrary to the original study, we found no elicitation of an AP negativity in any of the groups at electrode P7/T5 (bottom-left panel; gray-shaded area). Consistent with the original study, there were no group differences in the other temporal electrodes (bottom-right panel and upper panels; gray-shaded areas).

Frontal Negative Slow Wave (fnSW)

The two-way ANOVA of the fnSW amplitude at electrode location Fpz (calculated as the mean of electrodes Fp1 and Fp2) did not reveal a Group x Condition interaction ($F(2,101) = 0.92$, $p = 0.40$, $BF_{01} = 5.10$, $\eta^2_G = 0.003$). Furthermore, the three-way ANOVA using all electrodes did not reveal a Group x Condition x Electrode interaction ($F(5.92,298.82) = 1.77$, $p = 0.11$, $\eta^2_G = 0.002$). Note that for this analysis, we only used 19 electrodes because the mean amplitude of electrodes Fp1 and Fp2 (corresponding to electrode Fpz) was used instead of these electrodes themselves.

2.4.3 Electrophysiological Correlates of Absolute Pitch

Figure 4b visualizes the group-averaged waveforms in the Listening condition at medial temporal and posterior temporal electrodes. Itoh et al. found a High-AP group-specific temporal negativity at left-sided electrode P7/T5 with a latency of 150 ms (AP negativity). In this study, the waveforms did not clearly show an elicitation of an AP negativity in any of the groups (see Figure 4b, bottom-left panel).

Left Posterior Temporal Negativity (AP Negativity)

The one-way ANOVA of the AP negativity at electrode P7/T5 in the Listening condition did not reveal a main effect of Group ($F(2,101) = 0.04$, $p = 0.96$, $BF_{01} = 10.61$, $\eta^2_G < 0.001$). As shown in Figure 5, the component was not elicited in any of the three groups: Low-AP ($t(38) = -0.38$, $p = 0.71$, $BF_{01} = 5.42$, $d = 0.06$), Mid-AP ($t(39) = -0.63$, $p = 0.53$, $BF_{01} = 4.86$, $d = 0.09$), and, most importantly, High-AP ($t(24) = -0.11$, $p = 0.91$, $BF_{01} = 4.72$, $d = 0.02$).

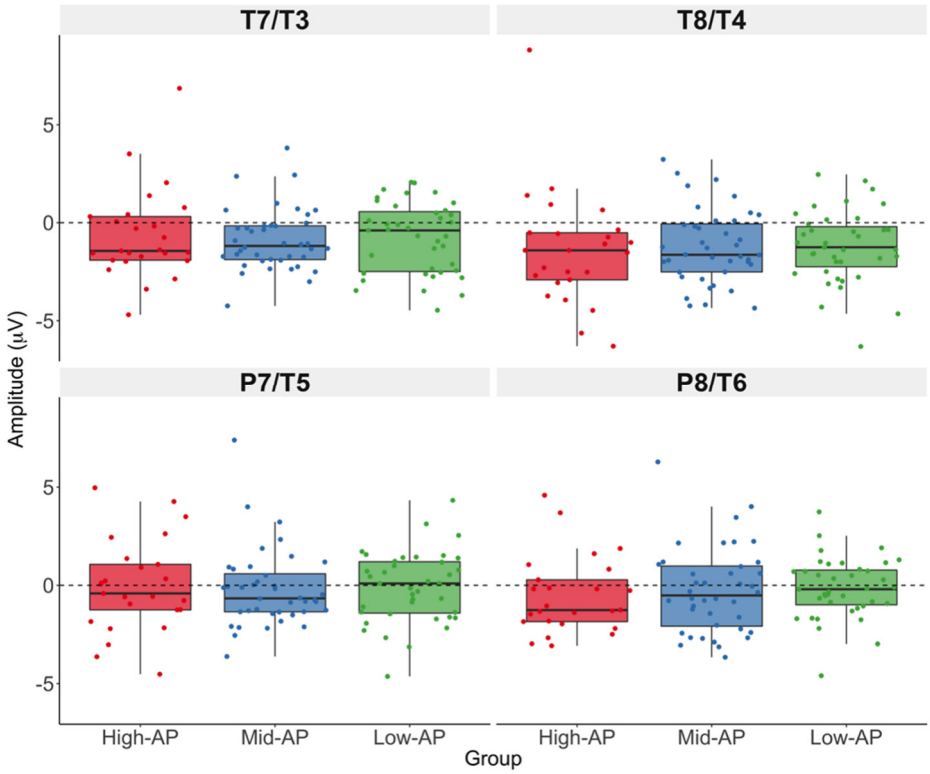


Figure 5. Lack of AP negativity elicitation.

We could not replicate the main finding of the original study — the elicitation of an AP negativity in High-AP musicians in the Listening condition at electrode P7/T5 (bottom-left panel). Furthermore, there were no group differences at other temporal electrodes during the AP negativity-specific time window (bottom-right panel and upper panel). This is consistent with the original study, in which the AP negativity showed a high topographic specificity and was not elicited at electrodes T7/T3, T8/T4, and P8/T6.

AP Negativity – Control Analysis

To ensure that the diverging results regarding the AP negativity were not caused by the slightly different EEG preprocessing in this study, we reran the complete analysis using a preprocessing pipeline which closely reproduced the preprocessing steps undertaken in the original study. To this end, we re-referenced the raw EEG data to the linked mastoids, before correcting ocular artifacts using independent component analysis. Then, we segmented and baseline-corrected the data. Segments containing artifacts were rejected using a threshold of ±100 µV, and the clean segments were averaged to compute ERPs. Finally, we low-pass filtered the ERPs at 30 Hz (48 dB/octave). The statistical analysis of the AP negativity yielded virtually identical results; the one-way ANOVA at P7/T5 did not reveal a main effect of Group ($F(2,101) = 0.006$, $p = 0.99$, $BF_{01} = 10.93$, $\eta^2_G < 0.001$). Thus, the results can be considered robust across different preprocessing pipelines.

2.5 Discussion

The primary goal of this study was to evaluate the presence of an AP negativity in High-AP musicians, an important finding in AP research originally reported by Itoh et al. (2005). Using a large sample size, we failed to replicate the elicitation of this AP-specific ERP component during tone listening. Bayesian inference in the form of Bayes factors indicated moderate evidence ($BF_{01} > 3$) in favor of the null hypothesis that no AP negativity was elicited in High-AP musicians (or in any of the other groups). We also found strong evidence in favor of the null hypothesis of no group differences in the AP negativity ($BF_{01} > 10$). The evaluation of the ERP correlates of RP yielded mixed results. We were able to replicate the group differences in the P3b at electrode Pz. Low-AP musicians showed a higher P3b than Mid-AP musicians, but there were no differences between Mid-AP and High-AP musicians. Thus, different from the original study, there was no gradual increase in P3b amplitude with lower levels of AP. Also, in contrast to the original study, the P3b was elicited in all groups including High-AP musicians. An identical pattern of results was found at Fz. With regard to the ppSW, we replicated the group differences at electrode Pz. Again, Low-AP musicians showed a higher ppSW than Mid-AP musicians, but there were no differences between Mid-AP and High-AP musicians. Unlike in the original study, all groups elicited a ppSW including High-AP musicians. Finally, we were not able to replicate the group differences in the fnSW. Bayes factors indicated moderate evidence for the null hypothesis of no fnSW group differences at electrode location Fpz ($BF_{01} > 3$).

Before turning to the detailed discussion of the findings of this replication study, it is crucial that we delineate the differences between this study and the original study. Several of these differences are effectively improvements we implemented to mitigate methodological issues of the original study. First and foremost, a critical difference between the two studies is that in Itoh et al.'s sample, the groups did not only differ in terms of AP proficiency, but also in musical experience. In fact, there was a statistically significant difference in years of musical training between Mid/High-AP musicians and Low-AP musicians; this difference makes it difficult to disentangle the differential effects of AP and musical experience as it has repeatedly been shown that musical training influences auditory evoked potential components (Fujioka, Ross, Kakigi, Pantev, & Trainor, 2006; Pantev et al., 1998; Schneider et al., 2002). In this replication study, the groups were matched with regard to musical experience (see Table 1). In a similar vein, none of the subjects in Itoh et al.'s sample were professional musicians, whereas our sample predominantly consisted of professional musicians and music students; this is also reflected in the substantial difference in years of musical training between Itoh et al.'s sample

(mean of all AP groups = 11.33 years) and our sample (mean = 19.82 years). Since the groups in this replication study were matched for years of musical training, the influence of this characteristic was controlled for and thus cannot account for the differential results. As the original study was conducted in Japan and this study was conducted in Switzerland, the two samples also differ with regard to their ethnicity. Furthermore, the sample of the original study predominantly consisted of female subjects, whereas the sex distribution in our sample was balanced (49% females). Nevertheless, in general, it is implicitly assumed that the conclusions drawn from the sample of a study are generalizable to the whole population of interest. In the original study, the authors thus made conclusions about the whole population of AP possessors and assumed that these conclusions would also hold for AP possessors irrespective of sex and different genetic and cultural background.

Moving on from the characteristics of the two samples, there are also differences regarding the statistical analyses that must be mentioned. In ERP research, the definition of the analyzed time window for a component is an essential analysis step. Itoh et al. defined the time windows based on a visual comparison of the group-specific waveforms. In these waveforms, the authors identified the components that differed between the groups and subsequently performed the statistical analysis on the time windows in which these differing components appeared. This procedure is an example for what Luck and Gaspelin (2017) call "multiple implicit comparisons", where the visual comparison of the group-specific waveforms correspond to a large number of statistical tests without an adjustment of the significance level. At the same time, this procedure makes the statistical analysis contingent on the results because the choice of the time window is not independent of the results. This is known as circular analysis which results in invalid inference and inflated effect sizes (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009; Vul, Harris, Winkielman, & Pashler, 2009). In cases of circular analysis, replications are crucial because they can provide a completely independent assessment of the extent to which the non-independent choice of the time windows had an influence on the results (Luck & Gaspelin, 2017). In this study, we thus used the time windows as defined in the original study. Another issue of the statistical analyses employed by Itoh et al. is the assumption of a difference between groups when one group showed a significant effect and another group showed a non-significant effect (Gelman & Stern, 2006; Nieuwenhuis et al., 2011). We thus performed statistical tests to directly compare the groups, in addition to the tests performed in the original study (more details are given in the Methods section).

As noted in the introduction, many studies cite Itoh et al.'s finding of an AP negativity at the left-sided electrode P7/T5 as evidence for the importance of the left planum temporale in AP,

not least because this interpretation was also given in the original study. This rationale is not compatible with the current standard of knowledge regarding the neuronal sources of scalp ERPs. Most importantly, ERPs at a particular electrode, for example the P7/T5 which is placed over the left posterior temporal cortex, are not simply generated by the neural activity in the patch of cortex directly underlying this electrode. On the contrary, the relationship between ERPs and its neuronal sources is extremely complex; even highly sophisticated algorithms are only able to provide rough estimations of these sources (Pascual-Marqui, Sekihara, Brandeis, & Michel, 2009). As Itoh et al. did not perform source estimation, the finding of an AP negativity in the original study should not be interpreted as evidence for the role of the left planum temporale in AP.

After having outlined the differences between the two studies, we now focus on the implications the results of this replication study have. How well does the non-elicitation of an early occurring AP-specific ERP component fit with evidence from previous EEG studies on AP? More than 25 years ago, Tervaniemi et al. (1993) investigated the mismatch negativity (MMN) in AP musicians. The MMN is an early ERP component elicited between 50 and 250 ms after stimulus onset and occurs in auditory oddball paradigms where a frequently presented standard stimulus is contrasted with an infrequently presented deviant stimulus. In their seminal study, Tervaniemi et al. (1993) found no differences in the MMN amplitude between musicians with and without AP that were matched in musical training. This finding was replicated in two independent studies (Greber, Rogenmoser, Elmer, & Jäncke, 2018; Rogenmoser et al., 2015). In a further study on the magnetic equivalent of the N1 component, Pantev et al. (1998) found differences between musicians and non-musicians, but again no differences between musicians with and without AP. The lack of differences in early auditory evoked potentials is also consistent with a more recent EEG study, in which there were no differences in the N1-P2 complex between AP musicians and Non-AP musicians (Elmer et al., 2013). Taken together, the non-elicitation of an early AP-specific component is compatible with a number of previous findings; a lack of early differences in auditory evoked potentials between musicians with and without AP as in this replication study seems to be the rule rather than the exception.

With regard to the ERP correlates of RP, the results of this replication study are in line with previous findings from EEG studies on AP. The reduced P3b component in subjects with higher levels of AP is not only consistent with Itoh et al.'s findings; it is also consistent with a number of EEG studies demonstrating a reduced or even absent P3b component in AP possessors during auditory oddball paradigms (Hantz et al., 1992; Klein et al., 1984; Wayman et al., 1992). A larger P3b amplitude is often associated with larger working memory demands (Polich, 2007).

Labeling tones using RP necessarily involves working memory processes because the currently perceived tone needs to be compared to previously perceived tones; from a (small) set of tones, the name of the tone can then be inferred by comparison after the whole set of tones was perceived at least once. Further evidence for the role of working memory in RP processing comes from functional imaging studies showing stronger activation in musicians without AP in brain regions such as the posterior part of the inferior frontal gyrus during the labeling of tones (Zatorre et al., 1998). The interpretation of lower ppSW amplitudes in subjects with higher levels of AP is less straightforward and the role of the ppSW in RP is unclear. Itoh et al. interpreted the ppSW component as a reflection of the process to associate the pitch stored in working memory and its label. At the same time, the authors provided little evidence to support this interpretation. As the ppSW component is highly dependent on the P3b (see waveforms in Figure 3), it is also possible that it reflected further working memory processing (García-Larrea & Cézanne-Bert, 1998). Future studies should try to elucidate the potential roles of the ppSW in RP. Lastly, in this replication study, the fnSW was not different between High-AP, Mid-AP, and Low-AP musicians. One reason for this non-replication could be that in the original study, this component was only elicited in the groups which had lower levels of musical training (Low-AP and Untrained). It is thus plausible that the elicitation of an fnSW only occurs in individuals with little to no musical training and that it might be unrelated to both RP and AP processing.

Since the very beginning of scientific AP research, there has been a discussion about the status of musicians who in tone-naming tests perform significantly better than chance-level but worse than highly proficient AP musicians (Bachem, 1955). These musicians have been termed *Quasi-AP* (Bachem, 1937) or *AP-2* (Baharloo et al., 1998; Loui, Li, Hohmann, & Schlaug, 2010; Loui et al., 2012). In both the original study and this replication study, these musicians most likely make up a large proportion of Mid-AP musicians. However, the categorization of subjects to subgroups according to certain cutoffs in tone-naming proficiency is completely arbitrary. Furthermore, it is unclear whether Mid-AP musicians are actually RP musicians using a strategy (e.g., using the memorized pitch of the tuning standard A4 in combination with RP processing to infer the pitch of a tone) or whether they are AP musicians performing badly in the tone-naming test for some unobservable reason (e.g., because they are lacking motivation). In that sense, it is crucial to additionally consult the self-report of the subjects as this can help distinguish AP from RP musicians. In this study, we repeatedly observed a lack of differences between High-AP and Mid-AP musicians. Additionally, Low-AP musicians performed only marginally above chance level in tone-naming and can thus be characterized as RP musicians (cf. Materials and Methods). Taken together, the findings of this replication study suggest that

a categorization into more than two groups, i.e. AP and RP musicians, may not be necessary. This notion of two separate populations for AP and RP musicians is also supported by large-scale behavioral studies employing tone-naming tests (Athos et al., 2007; Wengenroth et al., 2014).

A limitation of this study concerns the sample size. We invested considerable time and resources to recruit a large number of subjects (n = 104). However, due to the categorization of the subjects into three groups, the statistical power for some tests were slightly compromised. For example, even though descriptively there were differences at electrode Pz between High-AP and Low-AP musicians in both the P3b and the ppSW, these differences did not reach statistical significance and Bayes factors indicated inconclusive evidence (i.e. no hypothesis should be preferred). Inconclusive evidence, in turn, signifies the need for more subjects. As the recruitment of even more AP musicians is an extremely difficult task due to the rarity of the phenomenon, collaborations between multiple research groups should be considered to distribute the workload and foster the recruitment of even larger samples. A further limitation is that this study was not preregistered. Preregistration refers to the pre-specification of the hypotheses, the sample, the experimental procedure, and the statistical analyses of a study. These pre-specifications are then published on a website before the study is conducted. In general, preregistration protects against publication bias and analytic flexibility (Munafò et al., 2017). As we implemented some modifications to the pitch-naming task and, more importantly, to the statistical analyses, we could have preregistered these modifications to further increase the confidence in our findings.

To conclude, in this replication study using a large sample size, we were not able to replicate the elicitation of an AP negativity in High-AP musicians during tone listening, a landmark finding in neuroscientific AP research. We partially replicated the larger elicitation of the P3b and the ppSW, ERP correlates of RP, in musicians with lower levels of AP. As neither an isolated original study nor an isolated replication study can provide a final verdict on the veracity of a finding, more replication studies are imperative to further elucidate the neural underpinnings of AP.

Chapter 3 Empirical Work – Article 2

Neural patterns reveal single-trial information on absolute pitch and relative pitch perception

Simon Leipold, Marielle Greber, Silvano Sele, & Lutz Jäncke

3.1 Abstract

Pitch is a fundamental attribute of sounds and yet is not perceived equally by all humans. Absolute pitch (AP) musicians perceive, recognize, and name pitches in absolute terms, whereas relative pitch (RP) musicians, representing the large majority of musicians, perceive pitches in relation to other pitches. In this study, we used electroencephalography (EEG) to investigate the neural representations underlying tone listening and tone labeling in a large sample of musicians (n = 105). Participants performed a pitch-processing task with a listening and a labeling condition during EEG acquisition. Using a brain-decoding framework, we tested a prediction derived from both theoretical and empirical accounts of AP, namely that the representational similarity of listening and labeling is higher in AP musicians than in RP musicians. Consistent with the prediction, time-resolved single-trial EEG decoding revealed a higher representational similarity in AP musicians during late stages of pitch perception. Time-frequency-resolved EEG decoding further showed that the higher representational similarity was present in oscillations in the theta and beta frequency bands. Supplemental univariate analyses were less sensitive in detecting subtle group differences in the frequency domain. Taken together, the results suggest differences between AP and RP musicians in late pitch processing stages associated with cognition, rather than in early processing stages associated with perception.

This article was originally published in:

Leipold, S., Greber, M., Sele, S., & Jäncke, L. (2019). Neural patterns reveal single-trial information on absolute pitch and relative pitch perception. *NeuroImage*, *200*, 132-141. https://doi.org/10.1016/j.neuroimage.2019.06.030

3.2 Introduction

Music, speech, and many environmental sounds have acoustic waveforms that repeat over time. The repetition rate of these sounds is perceived as pitch (Plack et al., 2005). Pitch is not perceived equally by all humans: Patients with congenital amusia (tone deafness) are limited in judging whether a pitch is higher or lower than a previous pitch (Peretz, 2016). Healthy participants show a consistent preference to perceive either the fundamental pitch or the spectral pitch of a sound (Schneider et al., 2005). And then, there is a special group of musicians which effortlessly perceive, recognize, and name pitches in absolute terms and without the help of a reference pitch (Deutsch, 2013). These absolute pitch (AP) musicians stand in striking contrast to the large majority of relative pitch (RP) musicians which perceive pitches in relation to other pitches (McDermott & Oxenham, 2008; Miyazaki & Rakowski, 2002).

The neural correlates of AP and RP have been extensively studied using electroencephalography (EEG), yielding various findings regarding the auditory event-related potential (ERP) (Behroozmand, Ibrahim, Korzyukov, Robin, & Larson, 2014; Bischoff Renninger, Granot, & Donchin, 2003; Burkhard, Elmer, & Jäncke, 2019; Crummer, Walton, Wayman, Hantz, & Frisina, 1994; Elmer et al., 2013; Greber et al., 2018; Hantz et al., 1992; Itoh et al., 2005; Klein et al., 1984; Leipold, Oderbolz, Greber, & Jäncke, 2019; Rogenmoser et al., 2015; Tervaniemi et al., 1993; Wayman et al., 1992). All of these studies exclusively applied univariate analysis, which includes a preselection of electrodes and time windows to be analyzed. The classical univariate approach only utilizes a small portion of the collected data and disregards that the EEG signal is inherently multivariate (Michel & Murray, 2012). Furthermore, these studies calculated grand averages across numerous trials (Luck, 2014) and thus disregarded within-participant variance, which demonstrably contains information on perception and cognition (Ratcliff, Philiastides, & Sajda, 2009).

Recent methodological developments within the framework of "brain decoding" have enabled the successful extraction of information from single trials in spite of their low signal-to-noise ratio (Blankertz, Lemm, Treder, Haufe, & Müller, 2011; Grootswagers, Wardle, & Carlson, 2016). An increasing number of studies have used a decoding framework to analyze noninvasively obtainable electrophysiological signals (Cichy, Pantazis, & Oliva, 2014; Crouzet, Busch, & Ohla, 2015; King et al., 2013; Sankaran, Swaminathan, Micheyl, Kalluri, & Carlile, 2018; Schaefer, Farquhar, Blokland, Sadakata, & Desain, 2011). In the context of EEG decoding, it is useful to characterize the EEG signal as a time series of multivariate topographical patterns dynamically changing over time (Michel & Koenig, 2018). When transformed to the frequency domain, these topographical patterns become frequency-specific.

The patterns can be conceptualized as neural representations underlying the perceptual and cognitive processing that is performed in the context of experimental conditions. In other words, the patterns contain information about the experimental conditions, which can be decoded using multivariate pattern analysis (Grootswagers et al., 2016; Haxby, Connolly, & Guntupalli, 2014). To fully exploit the excellent temporal resolution of EEG, multivariate pattern analysis is best applied to separate time samples of the signal (Hausfeld, De Martino, Bonte, & Formisano, 2012). For each participant, a machine learning classifier is trained and tested (e.g., using cross-validation) at each time sample to differentiate between the experimental conditions (e.g., using the data from single trials). The success of this classification is quantified, for example, by the decoding accuracy (i.e. the fraction of trials that are correctly classified). This decoding accuracy in turn quantifies the representational (dis)similarity of the neural representations underlying the perceptual and cognitive processing in the different experimental conditions; the more similar the neural representations, the lower the decoding accuracy (Grootswagers et al., 2016; Kriegeskorte, Mur, & Bandettini, 2008).

In this study, we examined the neural representations underlying pitch perception in AP and RP musicians using time-resolved and time-frequency-resolved single-trial EEG decoding. Participants performed a pitch-processing task with two experimental conditions: a listening condition and a labeling condition. Cognitive models of AP suggest that pitch labeling is the most crucial component distinguishing AP from RP musicians (Levitin, 1994; Levitin & Rogers, 2005). More specifically, the label is part of the cognitive representation of the pitch and is automatically activated when a particular pitch is perceived (Levitin & Rogers, 2005). This automaticity has been empirically demonstrated using Stroop-like tasks, in which AP musicians responded slower in trials where the pitch of a tone did not match a simultaneously presented label (Akiva-Kabiri & Henik, 2012; Hsieh & Saberi, 2008b; Itoh et al., 2005; Schulze et al., 2013). Thus, we predicted that the neural representations underlying listening and labeling show high similarity in AP musicians because they automatically label pitches during listening and, to comply with the task demands, also label pitches during labeling. On the contrary, RP musicians do not label tones during listening, but they must use their RP ability during labeling to follow the instructions of the task. The labeling of pitches using RP involves additional cognitive processes not present during simple listening (Itoh et al., 2005; Zatorre et al., 1998). Thus, we predicted a lower representational similarity of listening and labeling in RP musicians. We tested these predictions using the EEG data of a large sample of musicians (n = 105) by comparing time series of representational similarity between AP musicians (n = 54) and RP musicians (n = 51). We also performed classical univariate analyses to maintain

comparability with previous studies. Finally, we employed a behavioral Stroop-like task to assess whether the postulated labeling automaticity is also present in our sample.

3.3 Materials and Methods

3.3.1 Participants

We analyzed the data of 105 participants, who were recruited in the context of a large project investigating AP (Brauchli, Leipold, & Jäncke, 2019; Burkhard et al., 2019; Greber et al., 2018; Leipold, Brauchli, Greber, & Jäncke, 2019). The EEG data of 104 of these participants was previously analyzed as part of a direct replication study using a different group assignment and methodology (Leipold, Oderbolz, et al., 2019). All participants were either professional musicians, music students, or highly trained amateurs, aged between 18 and 44 years. The participants' assignment to the groups of AP musicians or RP musicians was based on self-report. None of the participants reported any neurological, audiological, or severe psychiatric disorders. We confirmed the absence of hearing loss in all participants using pure tone audiometry (ST20, MAICO Diagnostics, Berlin, Germany). The demographical data (sex, age, handedness) and part of the behavioral data (tone-naming proficiency, musical aptitude, and musical experience) were collected using an online survey tool (http://www.limesurvey.org/). We used a German translation of the Annett questionnaire to verify the self-reported handedness (Annett, 1970). Musical aptitude was assessed using the Advanced Measures of Music Audiation (AMMA) (Gordon, 1989). To probe the RP ability of the participants, we employed a musical interval identification test. Crystallized intelligence was estimated in the laboratory using the Mehrfachwahl-Wortschatz-Intelligenztest (MWT-B) (Lehrl, 2005), and fluid intelligence was estimated using the Kurztest für allgemeine Basisgrößen der Informationsverarbeitung (KAI) (Lehrl, Gallwitz, Blaha, & Fischer, 1991). The participants provided written informed consent and were paid for their participation. The study was approved by the local ethics committee (http://www.kek.zh.ch/) and conducted according to the principles defined in the Declaration of Helsinki.

3.3.2 Tone-Naming Test

We employed a tone-naming test to assess the tone-naming proficiency of the participants (Oechslin et al., 2010). Participants had to name both the chroma and the octave (e.g., E4) of 108 pure tones presented in a pseudorandomized order. The test included all tones from C3 to B5 (twelve-tone equal temperament tuning, A4 = 440 Hz). Each of the tones was presented three times. The tones had a duration of 500 ms and were masked with 2,000 ms Brownian noise presented immediately before and after each tone. The tone-naming score was calculated

as the percentage of correct chroma identifications without considering octave errors (Deutsch, 2013); the chance level was at 8.3%.

3.3.3 Stroop-like Task

To empirically establish the labeling automaticity in AP musicians, we applied a behavioral audio-visual Stroop-like task (Akiva-Kabiri & Henik, 2012; Hsieh & Saberi, 2008b; Itoh et al., 2005; Schulze et al., 2013; Stroop, 1935). The auditory stimuli consisted of five pure tones (duration = 500 ms, 10 ms linear fade-in; 50 ms linear fade-out), which were created using Audacity (version 2.1.2, http://www.audacityteam.org/). The tones corresponded to C4 (262 Hz), D4 (294 Hz), E4 (330 Hz), F4 (349 Hz), and G4 (392 Hz) in twelve-tone equal temperament tuning. The visual stimuli were comprised of the musical notations (quarter notes in treble clef) of the same tones. Participants simultaneously heard a sine tone and viewed a musical note on a computer screen. They were asked to identify the visually presented musical notes as fast and as accurately as possible by button press (C, D, E, F, or G) while ignoring the tones. The label of the tone and the name of the musical note were either congruent or incongruent. If AP musicians are unable to suppress labeling, they are expected to experience more cognitive interference in incongruent trials than RP musicians. This would be reflected by larger response time differences between congruent and incongruent trials in AP musicians.

3.3.4 Experimental Procedure

All stimuli, the stimulus presentation scripts, and the raw data are available online (https://dx.doi.org/10.17605/OSF.IO/7QXJS).

EEG Experiment

During EEG data acquisition, the participants performed a pitch-processing task with a listening and a labeling condition (Itoh et al., 2005). The stimuli encompassed three pure tones (262 Hz, 294 Hz, and 330 Hz), corresponding to C4, D4, and E4, which were created using Audacity. The tones had a duration of 350 ms (10 ms linear fade-in; 50 ms linear fade-out) and were presented at a sound pressure level of 75 dB via on-ear headphones (HD 25-1, Sennheiser, Wedemark, Germany) using Presentation software (version 18.1, www.neurobs.com). Both experimental conditions consisted of 180 trials presented in a randomized order; each of the pure tones was presented 60 times. In each trial, first, a pure tone was presented. This tone was followed by a jittered inter-stimulus interval (duration = 900–1,100 ms). Then, an auditory cue (10 ms of pink noise; linear fade-in = 2 ms, linear fade-out = 2 ms) was presented to indicate to the participants that they should respond by a key press. After the cue, there was a silent period until a response was given, followed by an inter-trial interval of 1,000 ms. During the task, a

screen in front of the participants showed a black fixation cross on a gray background. The conditions only differed in the instructions given to the participants: In the listening condition, the participants were instructed to listen to the tones and press a neutrally marked key in response to the auditory cue, irrespective of the chroma of the tones. In the labeling condition, they were instructed to label the tones by pressing one of three corresponding keys marked with the tone names (C, D, and E). By instructing to respond only after the cue, we avoided a contamination of the EEG by motor artifacts. In both conditions, the participants were also instructed to respond as quickly and as accurately as possible and not to respond verbally. The order of the conditions was fixed across participants, with the listening condition always preceding the labeling condition. The whole task had a duration of 20 minutes.

EEG Data Acquisition and Preprocessing

The EEG was continuously recorded using an electrode cap (Easycap, Herrsching, Germany) with 32 Ag / AgCl electrodes arranged according to an extended 10/20 system (Fp1, Fp2, F7, F3, Fz, F4, F8, FT7, FC3, FCz, FC4, FT8, T7, C3, Cz, C4, T8, TP9, TP7, CP3, CPz, CP4, TP8, TP10, P7, P3, Pz, P4, P8, O1, Oz, O2) in combination with a BrainAmp amplifier (Brain Products, Munich, Germany). The sampling rate was 1,000 Hz, the data was online bandpass-filtered between 0.1 Hz and 100 Hz, the reference electrode was placed on the tip of the nose, and electrode impedance was kept below 10 kΩ throughout data acquisition by applying an electrically conductive gel.

The preprocessing of the EEG data was performed in BrainVision Analyzer (Version 2.1, https://www.brainproducts.com/). We used slightly different settings for the time-resolved EEG decoding and the time-frequency-resolved EEG decoding. In the preprocessing for the time-resolved EEG decoding, first, we bandpass-filtered the data from 0.5 Hz to 20 Hz (48 dB/octave) and applied a notch filter of 50 Hz. Next, we corrected vertical and horizontal eye movement artifacts using independent component analysis (Jung et al., 2000). Then, we removed the remaining artifacts using an automatic raw data inspection (removal criteria: amplitude gradient > 50 μV/ms, amplitude difference > 100 μV, amplitude minimum/maximum > –100 μV / +100 μV). Finally, we segmented the continuous data into epochs of 900 ms (–100 ms to +800 ms relative to stimulus onset) and applied a baseline correction using the time interval between -100 and 0 ms relative to stimulus onset. In the preprocessing for the time-frequency-resolved EEG decoding, we applied a high-pass filter of 0.5 Hz (48 dB/octave) and a notch filter of 50 Hz, but we did not use a low-pass filter (M. X. Cohen, 2014). The artifact correction procedure was identical to the time-domain analysis, but we used more liberal removal criteria during the automatic raw data inspection (amplitude

gradient > 100 µV/ms, amplitude difference > 200 µV, amplitude minimum/maximum > –200 µV / +200 µV). Lastly, we segmented the data into epochs of 2,500 ms (–1,000 ms to +1,500 ms relative to stimulus onset) and normalized the time-frequency power values (dB transformation) using the time interval between –500 and –100 relative to stimulus onset as a baseline. The prolonged epoch size avoided a contamination of the time-frequency representation by edge artifacts (M. X. Cohen, 2014).

3.3.5 Statistical Analysis of the Behavioral Data

The statistical analysis of the behavioral data was performed in R (version 3.3.2, https://www.r-project.org/). The participant characteristics were compared between the groups using Welch's t-tests. For the Stroop-like task, we analyzed the response times of the correct trials; trials with a response time shorter or longer than 2 standard deviations of the participant-and-condition-specific mean response time were excluded from the analysis. To calculate the size of the Stroop-effect for each participant, we subtracted the response times of the congruent trials from the response times of the incongruent trials. The differences in response times (incongruent minus congruent) were then subjected to a group comparison using Welch's t-tests. Effect sizes in the context of t-tests are given using Cohen's *d*. The significance level was set to $\alpha = 0.05$ for all behavioral analyses.

3.3.6 Time-resolved Single-Trial EEG Decoding

The goal of the EEG decoding analyses was to investigate the similarity of the dynamic neural representations underlying tone listening and tone labeling in AP musicians compared to RP musicians. As part of the time-resolved single-trial EEG decoding, we trained a machine learning classifier to differentiate between the experimental conditions at each time sample of the time-domain EEG signal. As a result, we obtained a dynamic measure of representational (dis)similarity operationalized by the time series of decoding accuracies per participant. The decoding accuracy represents the information available in an EEG topography at a given time sample, in our case, about the differences of neural representations underlying the experimental conditions (Grootswagers et al., 2016); similar representations result in lower decoding accuracies because the classifier confuses them more often than dissimilar representations. As EEG topographies rapidly change over tens to hundreds of milliseconds (Michel & Koenig, 2018), the available information simultaneously changes, leading to changes in decoding accuracy over time. It is not unusual that even before a stimulus is fully presented, the decoding accuracy increases as gradually more information becomes available (see e.g., Crouzet et al.,

2015). The resulting time series of decoding accuracies were statistically compared between AP and RP musicians using cluster-based permutation testing (Maris & Oostenveld, 2007).

In detail, for each participant, we resampled the preprocessed and segmented EEG data to 100 Hz using FieldTrip (version 20170713, http://www.fieldtriptoolbox.org/) in MATLAB R2016a. Then, built-in MATLAB functions were used to split and reshape the data to obtain a 32-dimensional vector of amplitudes per trial, per time sample, and per condition. Next, each vector was associated with its corresponding target (listening, labeling). The time sample-wise EEG decoding of the single trials was performed using scikit-learn (version 0.19.2, https://scikit-learn.org/) in Python 3.7.0. Separately for each time sample, we first z-transformed the dataset per feature (i.e. electrode) and subsequently performed the classification of the trials into listening and labeling using a linear Support Vector Machine (C = 1). As some participants had an unequal number of trials in the listening and the labeling condition due to the raw data inspection, we implemented the "balanced" mode of the classifier, in which the weights of the target classes are automatically adjusted to the class frequencies in the data. Decoding accuracy was assessed in a repeated 5-fold stratified cross-validation (100 iterations). Within one iteration, the trials were divided into five folds of approximately equal size, with the restriction that in each fold, the fraction of both target classes was representative of the whole dataset. The classifier was trained on four folds and the remaining fold was used for testing. This procedure was repeated until all folds had been used for testing once.

For the statistical group comparison, the resulting participant-wise time series of decoding accuracies were subsequently analyzed using cluster-based permutation testing in R. First, we calculated a time sample by time sample one-tailed Welch's t-test to compare decoding accuracies between the groups. Then, a cluster-defining threshold (CDT) of $p \leq 0.05$ was applied to build clusters of adjacent time samples in which decoding accuracies differed between the groups. Note that this CDT p value is used as a descriptive measure and not for inference. The size of the empirical clusters was used as the test statistic. This test statistic was compared to a null distribution of maximum cluster sizes that was obtained by repeating the procedure 10,000 times with permuted group labels. To account for temporal dependencies, we did not permute the group labels of single time samples within a time series of accuracies, but the group labels of every time sample belonging to the same time series were permuted as a whole. This is important because the temporal dependencies within a time series should be preserved, as otherwise the null distribution of cluster sizes is too liberal. The p value of the empirical clusters was defined as the fraction of cluster sizes obtained from the permuted data that were larger than the empirical cluster size. The significance level was set to $\alpha = 0.05$ at

cluster level. Please note that the cluster-based permutation test controls the family-wise error rate at the specified α at cluster level.

Post-hoc Source Estimation

To estimate the putative cortical sources underlying differences in time-resolved EEG decoding, we performed source modeling of the EEG signal using Brainstorm (version 3.4; https://neuroimage.usc.edu/brainstorm/) (Tadel, Baillet, Mosher, Pantazis, & Leahy, 2011). We used default Brainstorm settings unless otherwise stated. First, we computed a head model based on the default ICBM152 anatomical template with 15,002 dipole locations using the OpenMEEG boundary element method (Gramfort, Papadopoulo, Olivi, & Clerc, 2010) and calculated the noise covariance matrix based on the baseline (-100–0 ms) of the single trials. Then, we estimated distributed sources for each participant and condition using minimum norm estimation with loosely constrained dipole orientations (1 normal to cortex dipole, 2 tangential dipoles with amplitude = 0.2).

3.3.7 Time-Frequency-resolved Single-Trial EEG Decoding

The time-frequency-resolved EEG decoding was in large parts identical to the time-resolved EEG decoding. The crucial difference was the transformation of the time-domain signal to the frequency domain, which enabled frequency-band-specific EEG decoding to study the neural representations underlying tone listening and labeling. In detail, we first applied time-frequency analysis to the preprocessed EEG data to calculate the frequency-specific power over time (M. X. Cohen, 2014). The time-frequency analysis was performed using the FieldTrip method *mtmconvol* in combination with a single Hanning taper. The power values were calculated in a sliding time window of 500 ms (fixed length over frequencies) which was moved in 50 ms steps. For fixed window lengths, the frequency resolution (in Hz) is determined by the inverse of the window length (in s); this resulted in a frequency resolution of 2 Hz. We excluded frequencies below 4 Hz and above 30 Hz from the analysis because for frequencies in the delta range (below 4Hz), a 500 ms time window includes only one cycle, and for frequencies in the gamma range (above 30 Hz), there is a high possibility for contamination by muscle artifacts (Whitham et al., 2007). Consequently, we analyzed the following frequency bands: theta = 4–6 Hz, alpha = 8–12 Hz, and beta = 14–30 Hz. We then used built-in MATLAB functions to split and reshape the data to obtain a vector of power values per trial, per time sample, per condition, and per frequency band (theta, alpha, beta). We again associated each vector with its target (listening, labeling) and, separately for each frequency band, performed the time sample-wise single-trial EEG decoding using scikit-learn. This resulted in participant-wise time series

of decoding accuracies, which were subjected to cluster-based permutation testing in R, separately for each frequency band. See above for details concerning the EEG decoding and permutation testing.

3.3.8 Univariate Analysis of the EEG Data

To keep the study comparable with previous EEG studies on AP and RP, we additionally performed univariate statistics on the EEG data in both the time domain and the frequency domain. For the univariate analysis in the time domain, we resampled the data to 100 Hz. We computed ERPs by averaging the single trials per participant and condition. To analyze the interaction between group and condition, we calculated difference waves (labeling minus listening) for each participant. Based on the grand average of these difference waves across both groups, we restricted the analysis to the time interval between 300 and 600 ms relative to stimulus onset (cf. Figure 4A). We compared the difference waves between the groups using cluster-based permutation testing across both time and electrodes in FieldTrip (10,000 permutations, independent samples t-test, CDT $p \leq 0.05$, $\alpha = 0.05$ at cluster level; minimum number of electrodes for cluster = 4); in case of a significant interaction, we performed follow-up analyses by repeating the group comparison within each condition separately, and appropriately adjusted the significance level to $\alpha = 0.025$. To compare the univariate analysis in the time domain with the time-resolved EEG decoding, we repeated the group comparisons without restricting the statistical analysis to a predefined time interval (analysis window = 0–800 ms).

An analogous procedure was employed for the univariate analysis in the frequency domain. We averaged the single-trial time-frequency power to compute an average time-frequency representation per participant and condition. To analyze the group x condition interaction, we calculated the participant-wise difference between the conditions (labeling minus listening) in time-frequency power. Based on the grand average of the differences in time-frequency power across all participants, we restricted the analysis to the theta frequency band and the time interval between 300 and 800 ms (cf. Figure 5A). Finally, we performed cluster-based permutation testing on the differences in time-frequency power across time, frequency, and electrodes in FieldTrip (10,000 permutations, independent samples t-test, CDT $p \leq 0.05$, $\alpha = 0.05$ at cluster level); in case of a significant interaction, we performed follow-up group comparisons within each condition separately ($\alpha = 0.025$). Again, for a comparison with the time-frequency-resolved EEG decoding, we repeated the group comparisons for each of the three frequency bands (theta, alpha, and beta) without restricting the statistical analysis to a predefined time interval.

3.4 Results

3.4.1 Behavioral Results

Descriptive statistics of the participants' demographical and behavioral characteristics are given in Table 1. Group comparison of these characteristics revealed that there was no statistically significant difference in age ($t(100.97) = 1.33$, $p = 0.19$, $d = 0.26$), age of onset of musical training ($t(102.42) = –1.20$, $p = 0.23$, $d = 0.23$), cumulative musical training ($t(99.71) = 1.43$, $p = 0.16$, $d = 0.28$), crystallized intelligence ($t(102.86) = –1.49$, $p = 0.14$, $d = 0.29$), and fluid intelligence ($t(100.82) = –1.54$, $p = 0.13$, $d = 0.30$). We found a significant but nevertheless small difference in musical aptitude as measured by the AMMA total score ($t(100.99) = 2.14$, $p = 0.03$, $d = 0.42$), which was driven by higher AMMA tonal scores in AP musicians ($t(100.61) = 2.44$, $p = 0.02$, $d = 0.48$); the AMMA rhythm scores did not significantly differ ($t(101.38) = 1.53$, $p = 0.13$, $d = 0.30$). As shown in the left panel of Figure 1, we found a substantially better tone-naming proficiency in AP musicians ($t(102.92) = 13.94$, $p < 2.2 \times 10^{-16}$, $d = 2.72$). In the Stroop-like task, we found a significantly larger Stroop-effect in AP musicians associated with a medium effect size ($t(102.65) = 2.78$, $p = 0.007$, $d = 0.54$), which confirmed the presence of labeling automaticity in our sample of AP musicians (see Figure 1, right panel). There was no significant group difference in the interval identification score ($t(86.53) = 1.18$, $p = 0.24$, $d = 0.23$).

3.4.2 Exploratory correlation analyses among behavioral characteristics

As an association between tone naming and interval identification has been reported previously (Dooley & Deutsch, 2011), we tested the presence of this association in our sample using Pearson correlations. Across the whole sample, tone-naming proficiency and interval-identification scores were significantly correlated ($r = 0.31$, $p = 0.002$), and this association was also present within the groups of AP musicians ($r = 0.32$, $p = 0.02$) and RP musicians ($r = 0.40$, $p = 0.004$). These correlations were of medium size according to Cohen (1992). We also correlated the tone-naming proficiency with the size of the Stroop-effect to check whether the precision of tone naming was associated with the automaticity of labeling. We found a correlation between tone-naming and Stroop-effect across the whole sample ($r = 0.24$, $p = 0.02$), but this correlation was likely driven by group differences in both measures as we found no significant correlations within the groups (AP: $r = –0.12$, $p = 0.37$; RP: $r = 0.22$, $p = 0.13$).

Table 1. Participant characteristics.

Continuous measures given as mean ± standard deviation.

	AP musicians	RP musicians
Number of participants	54	51
Sex (female / male)	27 / 27	24 / 27
Handedness (right / left / both)	47 / 4 / 3	46 / 4 / 1
Age	26.67 ± 5.49 years	25.37 ± 4.49 years
Musical aptitude (AMMA) – total	66.11 ± 6.31	63.35 ± 6.86
Musical aptitude (AMMA) – tonal	32.33 ± 3.75	30.45 ± 4.13
Musical aptitude (AMMA) – rhythm	33.78 ± 2.83	32.90 ± 3.03
Age of onset of musical training	5.93 ± 2.39	6.49 ± 2.44
Cumulative musical training	16,592.51 ± 12,245.67 hours	13,539.00 ± 9,609.84 hours
Crystallized intelligence (MWT-B)	27.69 ± 5.10	29.10 ± 4.64
Fluid intelligence (KAI)	123.41 ± 32.16	132.19 ± 26.16
Tone-naming score	76.41 ± 19.55 %	24.00 ± 18.97 %
Stroop-effect	39.32 ± 38.71 ms	18.32 ± 38.74 ms
Interval identification score	77.59 ± 15.94 %	73.00 ± 22.97 %

Abbreviations: AMMA = Advanced Measures of Music Audiation, AP = absolute pitch, KAI = Kurztest für allgemeine Basisgrößen der Informationsverarbeitung, MWT-B = Mehrfachwahl-Wortschatz-Intelligenztest, RP = relative pitch.

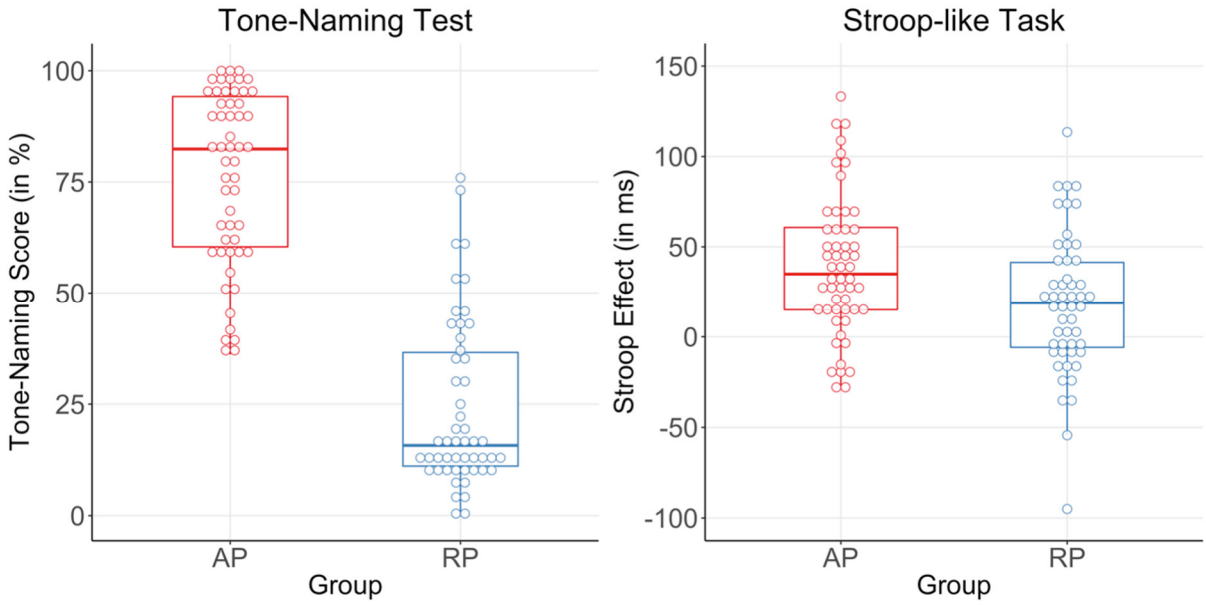


Figure 1. Results of the tone-naming test and the Stroop-like task.

The tone-naming test revealed a substantially better tone-naming proficiency in AP musicians than RP musicians (left panel). The Stroop-like task confirmed the presence of labeling automaticity in our sample as AP musicians showed a larger Stroop-effect than RP musicians (right panel). Abbreviations: AP = absolute pitch, RP = relative pitch.

3.4.3 Results of the Time-Resolved EEG Decoding

The participant-wise time-resolved EEG decoding of listening and labeling revealed that in the early stages of pitch perception, decoding accuracies were near chance level (i.e. the representational similarity of the conditions was high) in many participants of both groups. In later stages, decoding accuracies increased before becoming lower again during very late stages of pitch perception. See Supplementary Figure 1A for a participant-wise visualization of time-resolved EEG decoding and Supplementary Figure 1B for a participant-wise visualization of statistical significance of time-resolved EEG decoding. The group comparison of the decoding time series using cluster-based permutation testing yielded a significant cluster in the time interval between 430 and 550 ms after stimulus onset ($p = 0.01$, empirical cluster size $k = 13$). Consistent with our predictions, EEG decoding was significantly lower in AP musicians and thus, the neural representations underlying tone listening and labeling were more similar. During and immediately after stimulus presentation (~0–400 ms), we found no evidence for a difference in decoding accuracies between the groups. Figure 2A visualizes the group-wise decoding time series, and Figure 2B visualizes the null distribution of cluster sizes (obtained through 10,000 permutations) of the time-resolved analysis.

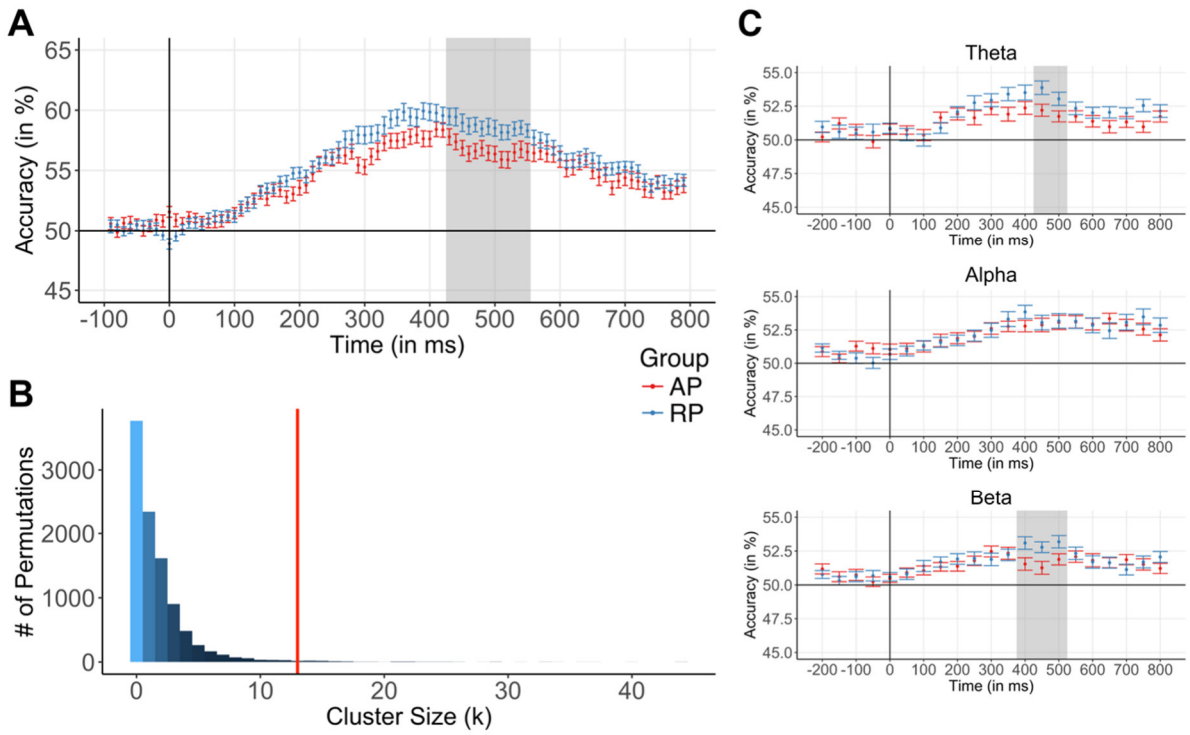


Figure 2. Single-trial EEG decoding in the time and frequency domain.

Group differences in the representational similarity of listening and labeling were found both in the time domain and the frequency domain. (A) Time-resolved EEG decoding revealed group differences in EEG decoding 430–550 ms after stimulus onset (gray shaded area). (B) The null distribution of cluster sizes was obtained by repeating the analysis 10,000 times with permuted group labels. The empirical cluster size (k = 13) is depicted by the red vertical line. (C) Time-frequency-resolved EEG decoding revealed group differences in representational similarity in the theta frequency band 450–500 ms after stimulus onset (gray shaded area in the upper panel) and the beta frequency band 400–500 ms after stimulus onset (gray shaded area in the lower panel). In the alpha frequency band, EEG decoding did not differ between the groups. Error bars represent the standard error of the mean per time sample and group. Abbreviations: AP = absolute pitch, RP = relative pitch.

Results of the Post-hoc Source Estimation

We performed source estimation for the time interval where we found significant group differences in representational similarity (430–550 ms). For each condition separately, we averaged the source-space current density values across time between 430 and 550 ms using the absolute values of the three dipole orientations (1 normal, 2 tangential). Subsequently, we subtracted the current density values in the listening condition from the values in the labeling condition and averaged these differences per group. As shown in Figure 3, in both groups, maximal condition differences in source space were located in the presupplementary motor area (preSMA), and, to a lesser degree, in the medial superior parietal cortex and the dorsolateral prefrontal cortex. Descriptively, condition differences in source space were stronger in RP musicians than AP musicians. We refrained from additional statistical inference in source space as this would constitute a circular analysis, given that we identified the time window based on significant group differences in time-resolved EEG decoding.

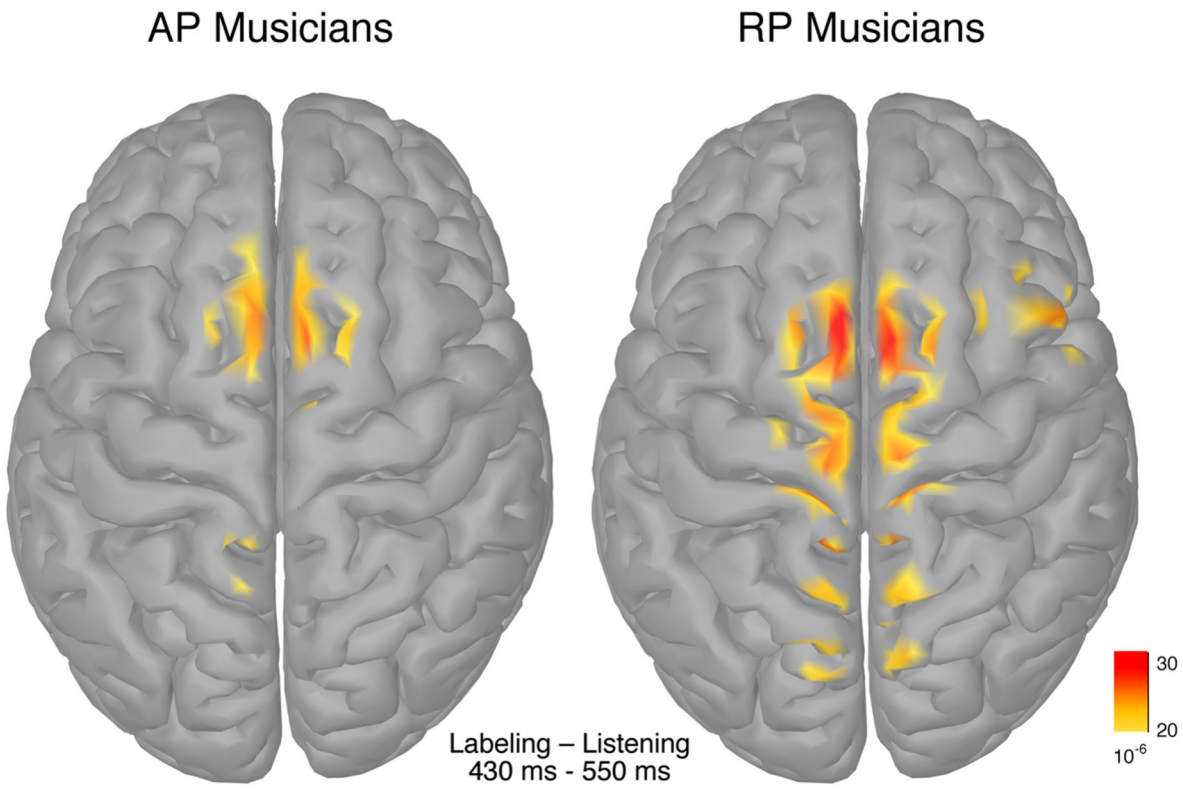


Figure 3. Group-specific condition differences in source space.

The group-specific condition differences (labeling minus listening) for the time interval where representational similarity of listening and labeling differed between the groups (430–550 ms) were located in the presupplementary motor area, and, to a lesser degree, in the medial superior parietal cortex and the dorsolateral prefrontal cortex. For visualization purposes, the images were thresholded at 60% of the maximal current density amplitude.

3.4.4 Results of the Time-Frequency-Resolved EEG Decoding

The participant-wise time-frequency-resolved EEG decoding of the experimental conditions revealed that in the theta and beta frequency bands, for many participants, decoding accuracies were relatively low during the first stages of pitch perception (i.e. the neural representations were similar), gradually increased during later stages and then dropped again in the very late stages (cf. Supplementary Figures 2A, 2B, 3A, and 3B). The group comparison of the time-frequency-resolved decoding time series using cluster-based permutation testing revealed significant group differences in EEG decoding in the theta and the beta frequency bands. Analogous to the time-resolved analysis, in the theta frequency band, we found significantly lower decoding accuracies in AP musicians between 450 and 500 ms after stimulus onset (p = 0.04, k = 2). In the beta frequency band, we found lower decoding accuracies in AP musicians between 400 and 500 ms (p = 0.02, k = 3). In contrast, we did not find evidence for differences in EEG decoding in the alpha frequency band. Figure 2C visualizes the group-wise decoding time series for each frequency band.

3.4.5 Results of the Univariate Analyses

Using a restricted analysis window, the univariate analysis of the time-domain EEG data revealed a group (AP vs. RP) x condition (listening vs. labeling) interaction 380–530 ms after stimulus onset (p = 0.02, k = 16), characterized by smaller difference wave amplitudes (labeling minus listening) in AP musicians at predominantly frontal, central, and parietal electrodes (see Figure 4B and Figure 4C). Post-hoc group comparisons separately within each condition revealed no significant amplitude differences (all $p > 0.025$). However, there was one cluster of lower amplitudes in AP musicians during labeling, which did not survive the cluster-based correction for multiple comparisons (p = 0.11, k = 3). From this it follows that in the time domain, the groups could primarily be differentiated based on their condition differences; the group difference within each condition, though, was too small to be reliably detected. The univariate time-domain analysis without a restriction of the temporal analysis window revealed the same interaction as the restricted analysis (p = 0.03, k = 16).

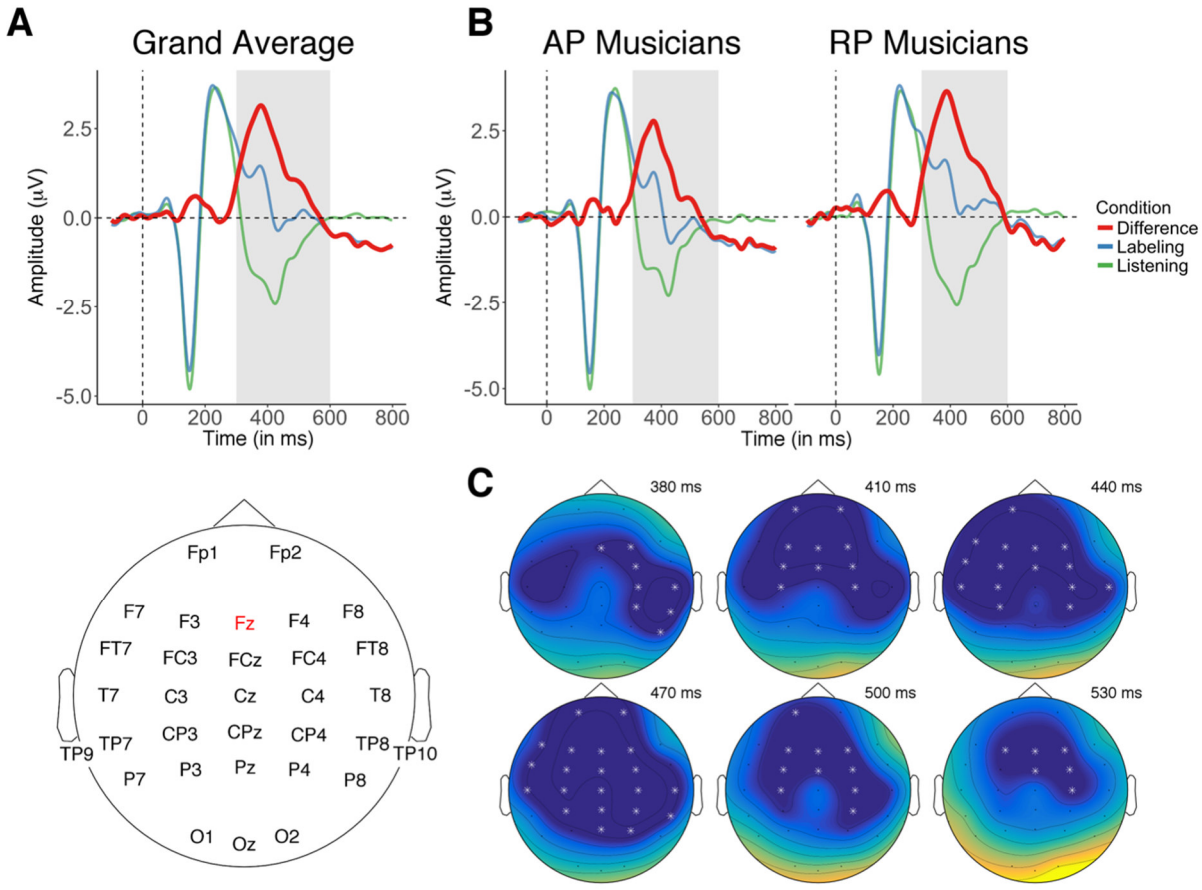


Figure 4. Univariate analysis of the time-domain EEG data

(A) Based on the grand-average difference wave of all participants (in red) at electrode Fz, we restricted the cluster-based permutation testing to 300–600 ms after stimulus onset (gray shaded area). We also analyzed the time-domain EEG data without this restriction (see main text). (B) Waveforms per condition (in blue and green) and difference waveform (in red), separately for AP musicians (left) and RP musicians (right). (C) Using a restricted analysis window, cluster-based permutation testing revealed reduced difference wave amplitudes (labeling minus listening) in AP musicians at predominantly frontal, central, and parietal electrodes. The highlighted electrodes showed the strongest group differences.

Using a restricted analysis window, the univariate analysis in the frequency domain revealed a significant group x condition interaction in the theta frequency band 450–750 ms after stimulus onset ($p = 0.04$, $k = 7$); this interaction was characterized by smaller power differences between conditions (labeling minus listening) in AP musicians (see Figure 5B). The post-hoc analyses separately within each condition revealed no group differences during listening but smaller theta power in AP musicians during labeling ($p = 0.003$, $k = 11$). The univariate frequency-domain analysis without restriction of the analysis window only yielded a trend towards a group x condition interaction ($p = 0.05$, $k = 7$). Furthermore, the univariate frequency-domain analysis did not reveal significant group differences in the alpha and beta frequency bands without a restriction of the analysis window.

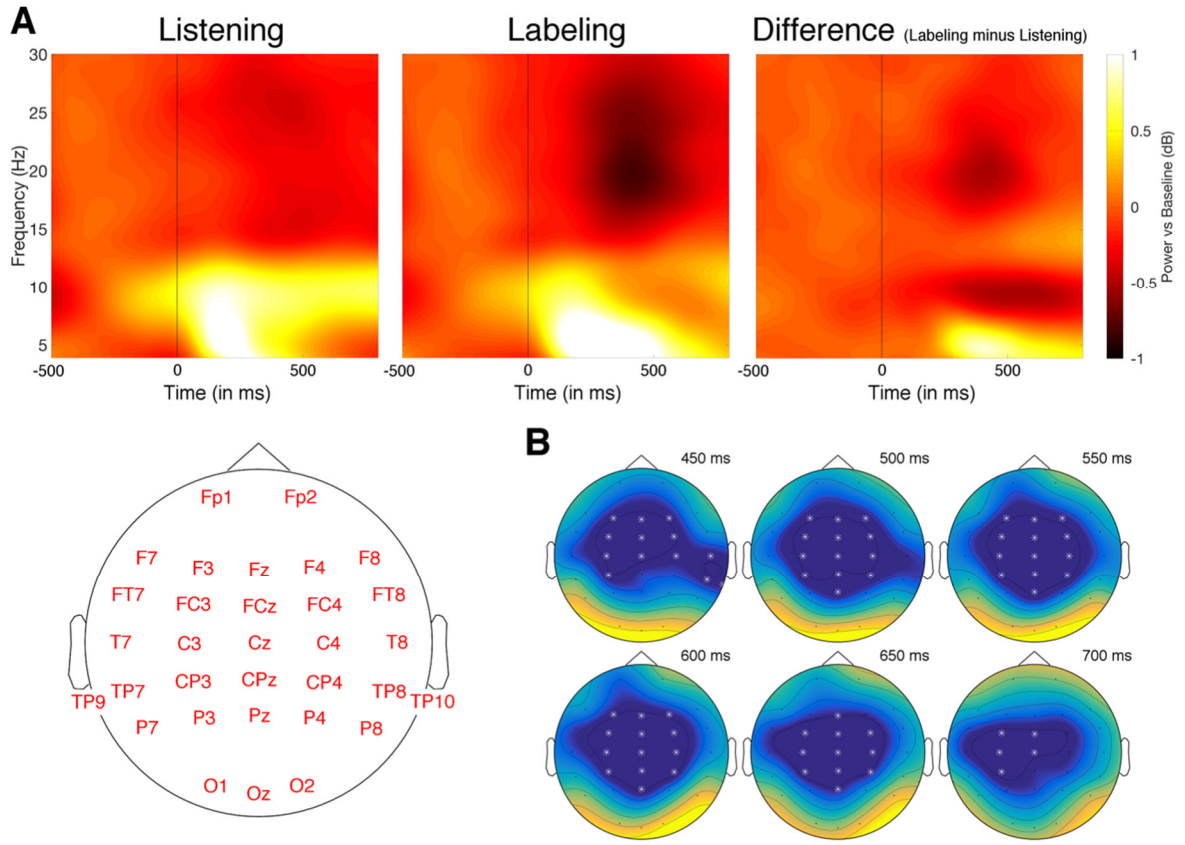


Figure 5. Univariate analysis of the frequency-domain EEG data

(A) Based on the grand average of differences in time-frequency power across all participants (right panel), we restricted the cluster-based permutation testing to the theta frequency band (4–6 Hz) and the time interval between 300 and 800 ms after stimulus onset. We also analyzed the frequency-domain EEG data without this restriction (see main text). The mean time-frequency power over all electrodes is visualized. (B) The restricted cluster-based permutation testing revealed reduced theta power differences (labeling minus listening) in AP musicians. The highlighted electrodes showed the strongest group differences in the theta frequency band.

3.5 Discussion

The conceptualization of multivariate EEG patterns as decodable neural representations is a powerful tool for investigating the dynamic neural mechanisms underlying perception and cognition (Grootswagers et al., 2016). As explained in the introduction, both cognitive models of AP and empirical evidence from previous studies suggest more similar representations of tone listening and labeling in AP musicians than in RP musicians. However, no study has explicitly tested this prediction. Here, we studied the neural representations underlying listening and labeling in a large sample of AP and RP musicians using a "brain-decoding" approach. As predicted, time-resolved single-trial EEG decoding revealed higher representational similarity of listening and labeling in AP musicians during late stages of pitch perception. Time-frequency-resolved single-trial EEG decoding revealed that the higher representational

similarity in AP musicians was present in oscillations in the theta and beta frequency bands. In the alpha frequency band, we found no evidence for group differences with regard to the similarity of neural representations underlying listening and labeling. Univariate analysis in the time domain revealed that AP musicians had lower amplitude differences between the conditions. Frequency-domain univariate analysis revealed lower theta power differences in AP musicians driven by lower theta during labeling. In contrast to the decoding approach, the univariate approach did not identify group differences in the beta frequency band.

The dynamics of representational similarity were characterized by a lack of differences between AP and RP musicians during the early stages of pitch perception; only after 300 ms, group differences began to emerge, and these differences were strongest during late processing stages. The early processing of tones encompasses the perceptual analysis of acoustic features of the stimulus (Koelsch & Siebel, 2005); hence, our results suggest that AP and RP musicians do not differ in this initial analysis of the tones. This is in accordance with a number of previous EEG studies not finding differences between AP and RP musicians in early ERPs (Elmer et al., 2013; Greber et al., 2018; Pantev et al., 1998; Rogenmoser et al., 2015; Tervaniemi et al., 1993). On the other hand, much evidence points to differences between AP and RP musicians in late ERPs associated with higher-order cognitive functions (Bischoff Renninger et al., 2003; Crummer et al., 1994; Elmer et al., 2013; Hantz et al., 1992; Klein et al., 1984; Wayman et al., 1992). We extend these studies by demonstrating that by using a decoding approach, group differences are detectable on a single-trial level. Consistent with the notion that late stages of pitch perception are associated with cognitive processing, the putative cortical sources underlying the group differences were located in the superior frontal and parietal cortices.

In previous EEG and neuroimaging studies, AP and RP have been repeatedly associated with tonal working memory, e.g., in the context of auditory oddball paradigms (Hantz et al., 1992; Klein et al., 1984). The original idea was that RP musicians need to update working memory when being confronted with novel incoming tones, whereas AP musicians do not rely on working memory processes during pitch perception because they have fixed long-term memory templates for tones (Klein et al., 1984). This idea has been further developed to suggest that AP musicians do not need to access working memory to name musical intervals (Hantz et al., 1992; Zatorre et al., 1998) or to label single tones (Itoh et al., 2005); the automaticity of tone labeling as demonstrated by our Stroop-like task underlines this observation. Furthermore, in recent behavioral studies, tone-naming proficiency has been linked to working memory capacity (Deutsch & Dooley, 2013; Van Hedger, Heald, Koch, & Nusbaum, 2015). Taken together, in the case of our pitch perception task, the dissimilarity of the neural representations underlying

listening and labeling in RP musicians might be due to them accessing working memory during labeling (but not during listening).

To date, only a single study has investigated the neural oscillations of AP and RP during musical perception (Behroozmand, Ibrahim, Korzyukov, Robin, & Larson, 2015). Using univariate analysis, the authors did not identify differences between AP and RP musicians. In this study, we found group differences in theta and beta oscillations in late stages of pitch perception. The exact role of neural oscillations in perception and cognition is still somewhat of a mystery (X.-J. Wang, 2010), but both theta and beta oscillations have been associated with various cognitive functions. In the auditory domain, it has been proposed that theta oscillations have a causal role in enhancing tonal working memory (Albouy, Weiss, Baillet, & Zatorre, 2017). Beta oscillations have been linked to top-down processing (Engel & Fries, 2010).

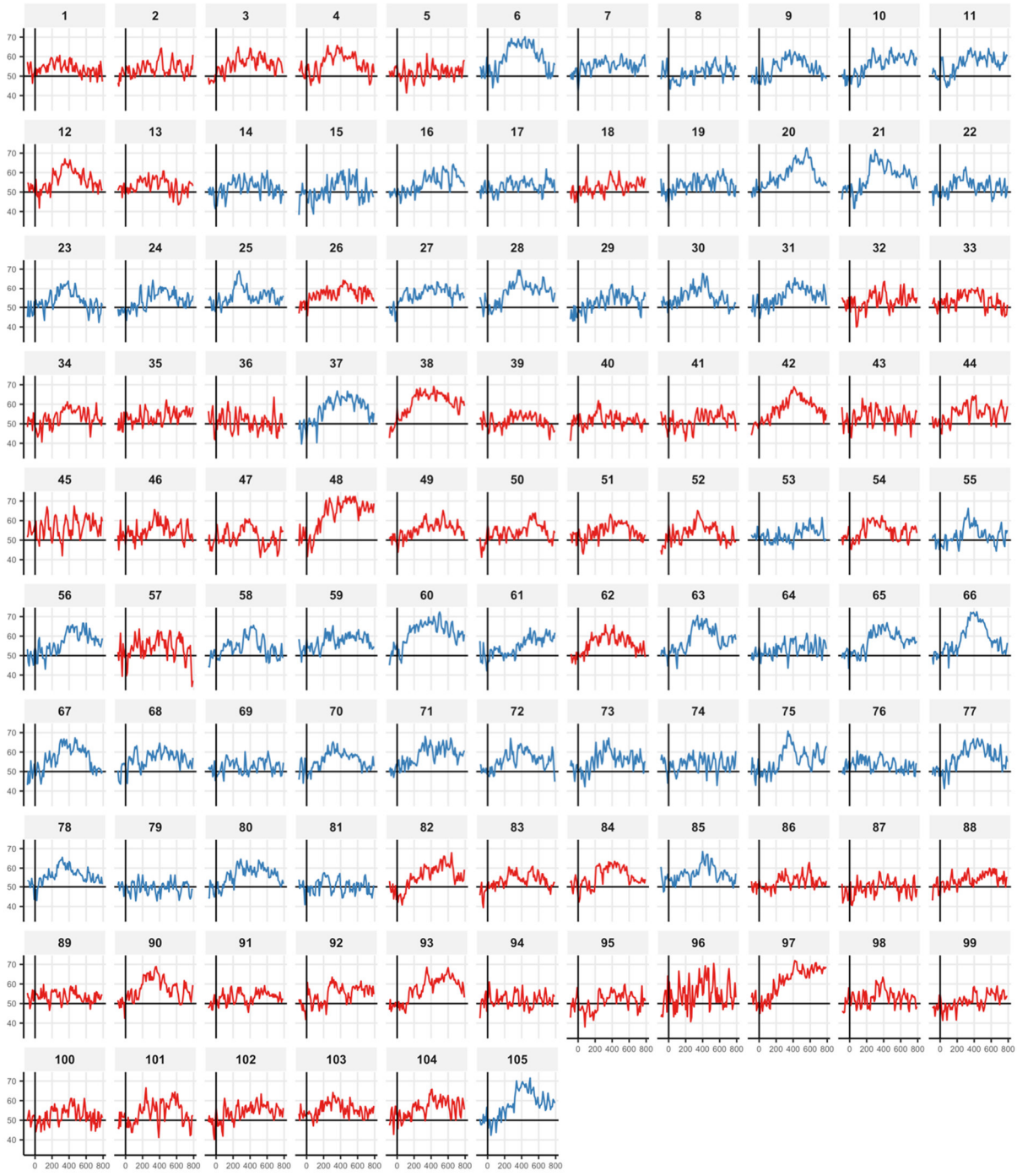
However, as we did not collect behavioral measurements of higher-order cognitive functions apart from the Stroop-like task, the specific nature of the cognitive functions underlying the group differences in late stages of pitch perception remains speculative. Future studies should include such behavioral measurements to better determine the underlying cognitive processing.

As a supplement to the EEG decoding approach, we used classical univariate analysis, first and foremost to make our study comparable with previous ERP studies on AP. We found, on average, lower amplitude differences and, using a restricted analysis window, lower theta power differences in AP musicians. These results are somewhat comparable to the results of the EEG decoding analysis. However, using the same analysis windows for both approaches, the decoding approach showed higher statistical sensitivity in the frequency domain as it exclusively identified a group difference in the beta frequency band. In addition, the univariate analysis without a restricted analysis window only revealed a trend towards lower theta power differences in AP musicians. This is consistent with observations regarding a higher sensitivity of multivariate compared to univariate methods in detecting subtle distributed effects (Grootswagers et al., 2016). Therefore, one should not mistakenly equate the results of the two analyses approaches; they presumably reflect two complementary aspects of the same underlying neural mechanisms, but the relationship of these two aspects is still subject to ongoing discussions (Jimura & Poldrack, 2012). In the case of EEG, a major advantage of (time-resolved) decoding is that it summarizes the information from multiple experimental conditions, electrodes, and trials into a single time series that can potentially be related to signals from other imaging modalities, e.g., neuroimaging data, behavioral data, or even the outputs of computational models (Kriegeskorte et al., 2008). In this context, future studies investigating the neural representations underlying pitch perception should try to sample more

than two conditions (listening, labeling) in the vast space of possible experimental conditions to fully exploit the capabilities of this approach. Finally, it should be noted that apart from representational similarity analysis as it was performed here, there exist other analysis methods that allow for a multivariate analysis of EEG patterns, e.g., the multivariate general linear model (Friston et al., 1996).

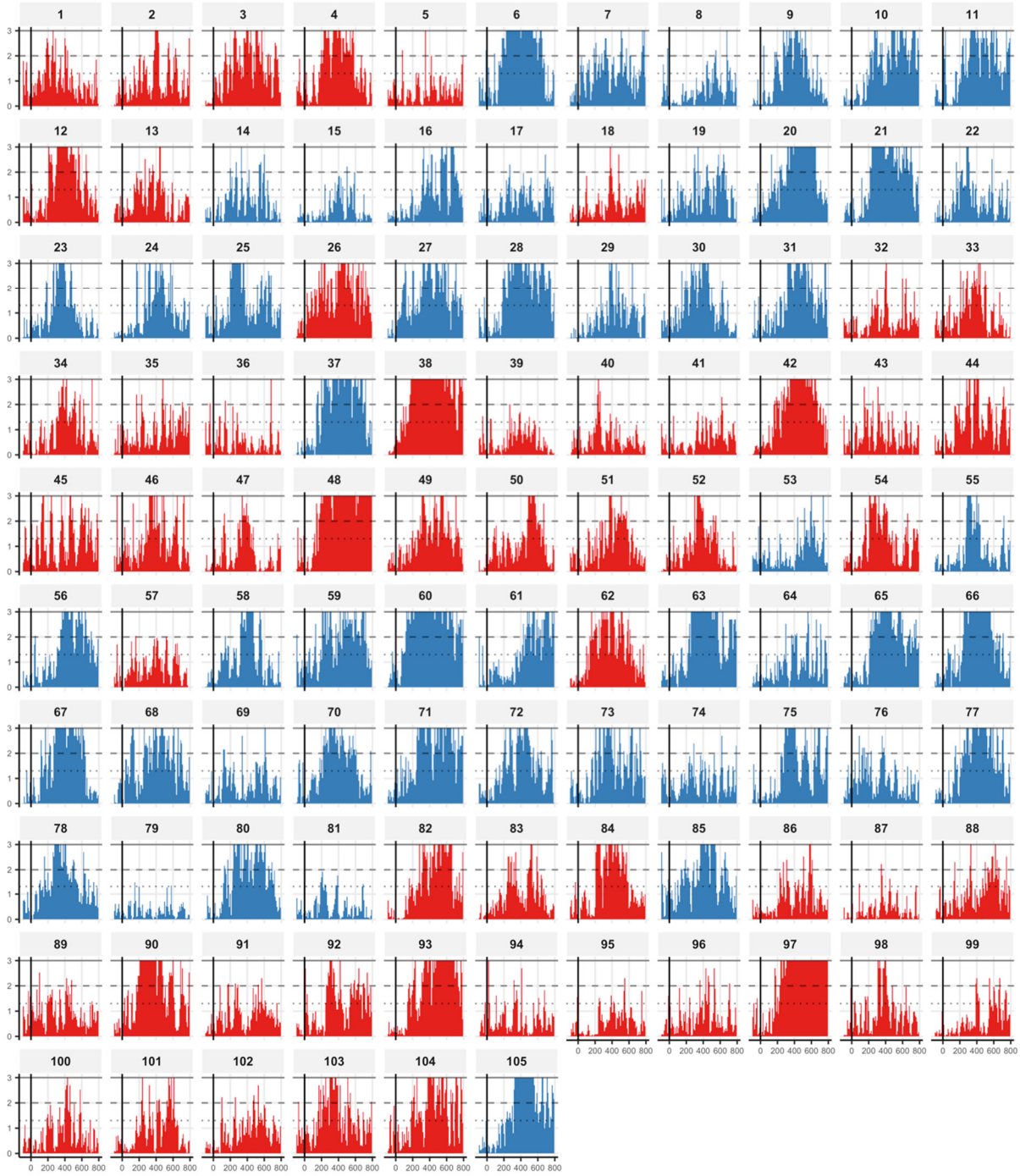
In conclusion, we showed that the neural representations underlying listening and labeling are more similar in AP musicians than RP musicians during late stages of pitch perception; this effect was present in oscillations in the theta and beta frequency bands. Compared to the novel decoding approach, the conventional univariate approach was less sensitive in identifying subtle group differences in the frequency domain.

3.6 Supplementary Data



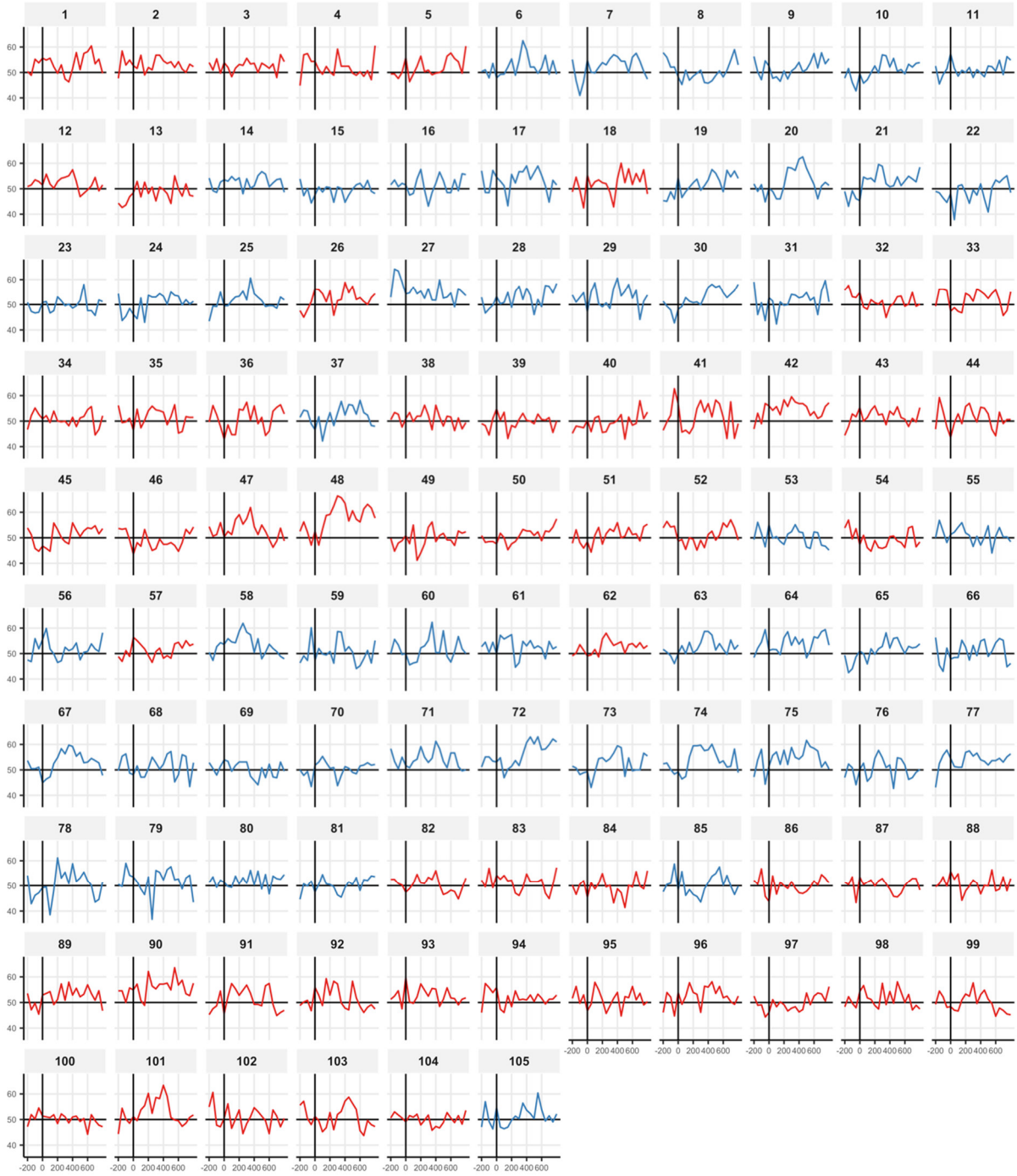
Supplementary Figure 1A. Participant-wise decoding time series of the time-resolved EEG decoding.

The x-axis represents time (in ms) relative to stimulus onset; the y-axis represents decoding accuracy (in %). The solid black vertical line denotes stimulus onset. The solid black horizontal line at y = 50 denotes chance level (50%). Absolute pitch musicians are colored in red; relative pitch musicians are colored in blue.



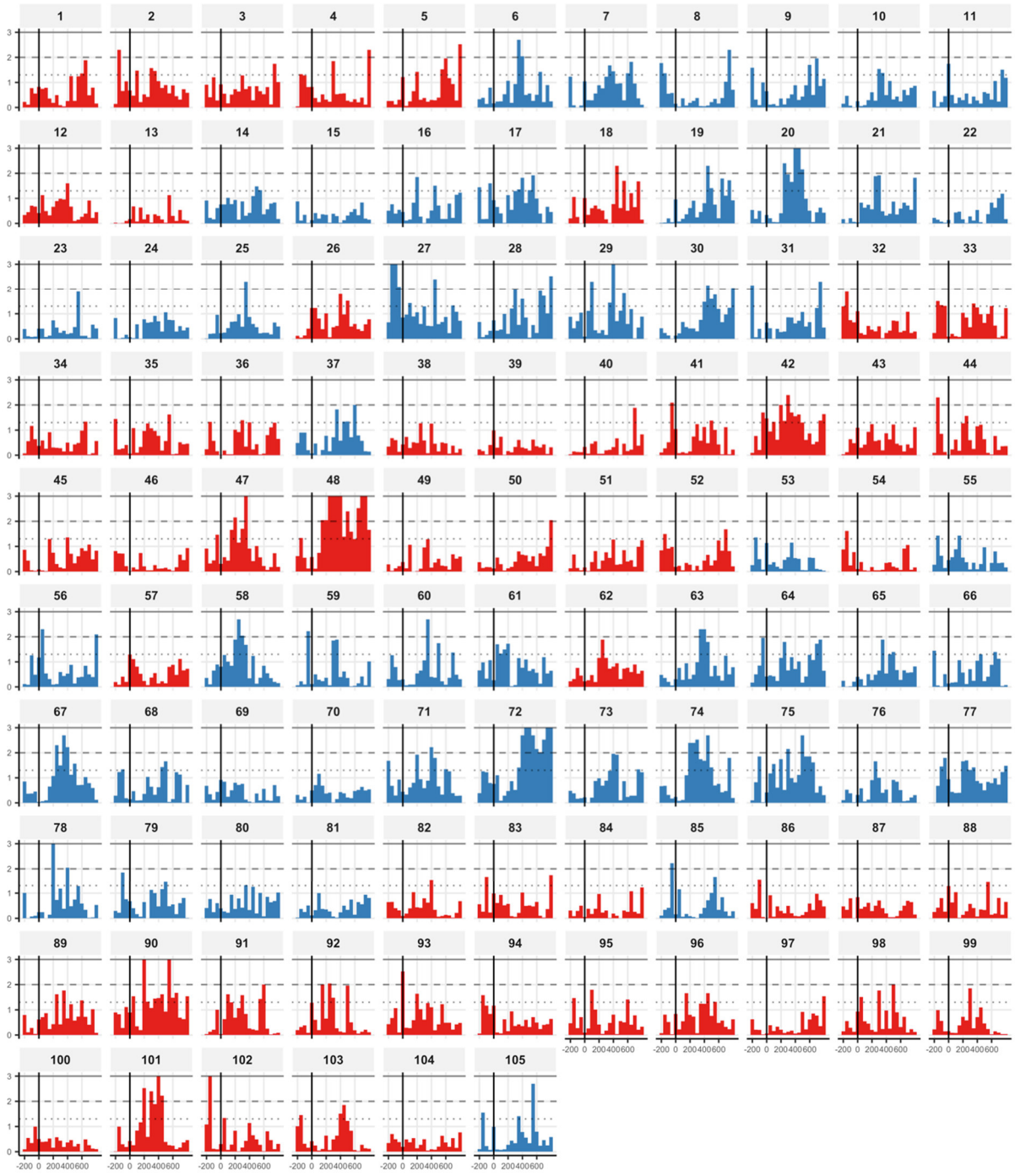
Supplementary Figure 1B. Participant-wise $-\log_{10} p$ value time series of the time-resolved EEG decoding.

Group differences in representational similarity (operationalized by decoding accuracy) are unlikely to be meaningful if the accuracies in both groups are below chance level (50%). Thus, we assessed the statistical significance of EEG decoding for each participant and time bin separately using 1,000 random permutations. The x-axis represents time (in ms) relative to stimulus onset; the y-axis represents $-\log_{10} p$ values. The solid black vertical line denotes stimulus onset. The dotted horizontal line at y ≈ 1.3 denotes $p = 0.05$, the dashed horizontal line at y = 2 denotes $p = 0.01$, and the solid horizontal line at y = 3 denotes $p = 0.001$. Absolute pitch musicians are colored in red; relative pitch musicians are colored in blue.



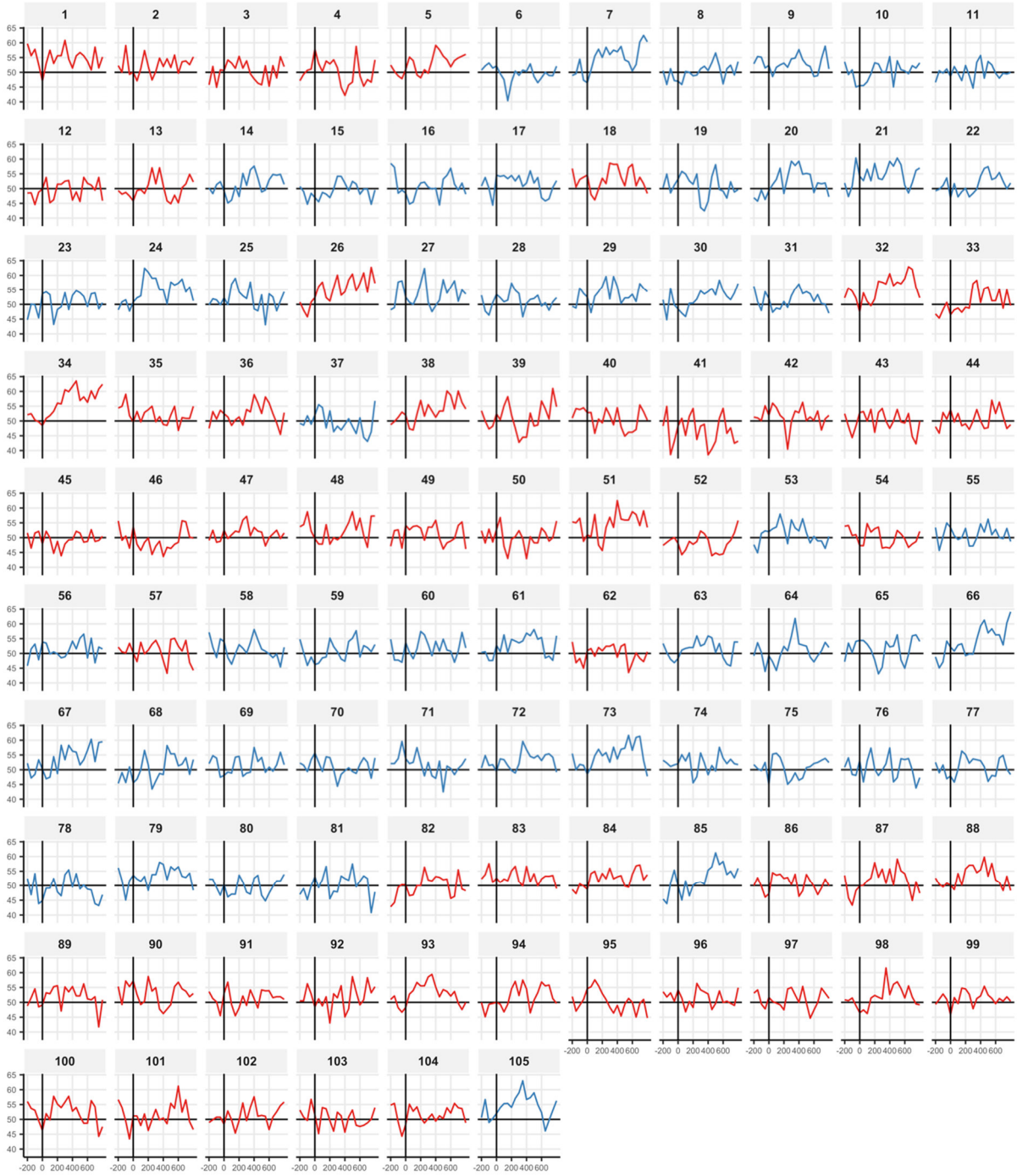
Supplementary Figure 2A. Participant-wise decoding time series of the time-frequency-resolved EEG decoding in the theta frequency band.

The x-axis represents time (in ms) relative to stimulus onset; the y-axis represents decoding accuracy (in %). The solid black vertical line denotes stimulus onset. The solid black horizontal line at y = 50 denotes chance level (50%). Absolute pitch musicians are colored in red; relative pitch musicians are colored in blue.



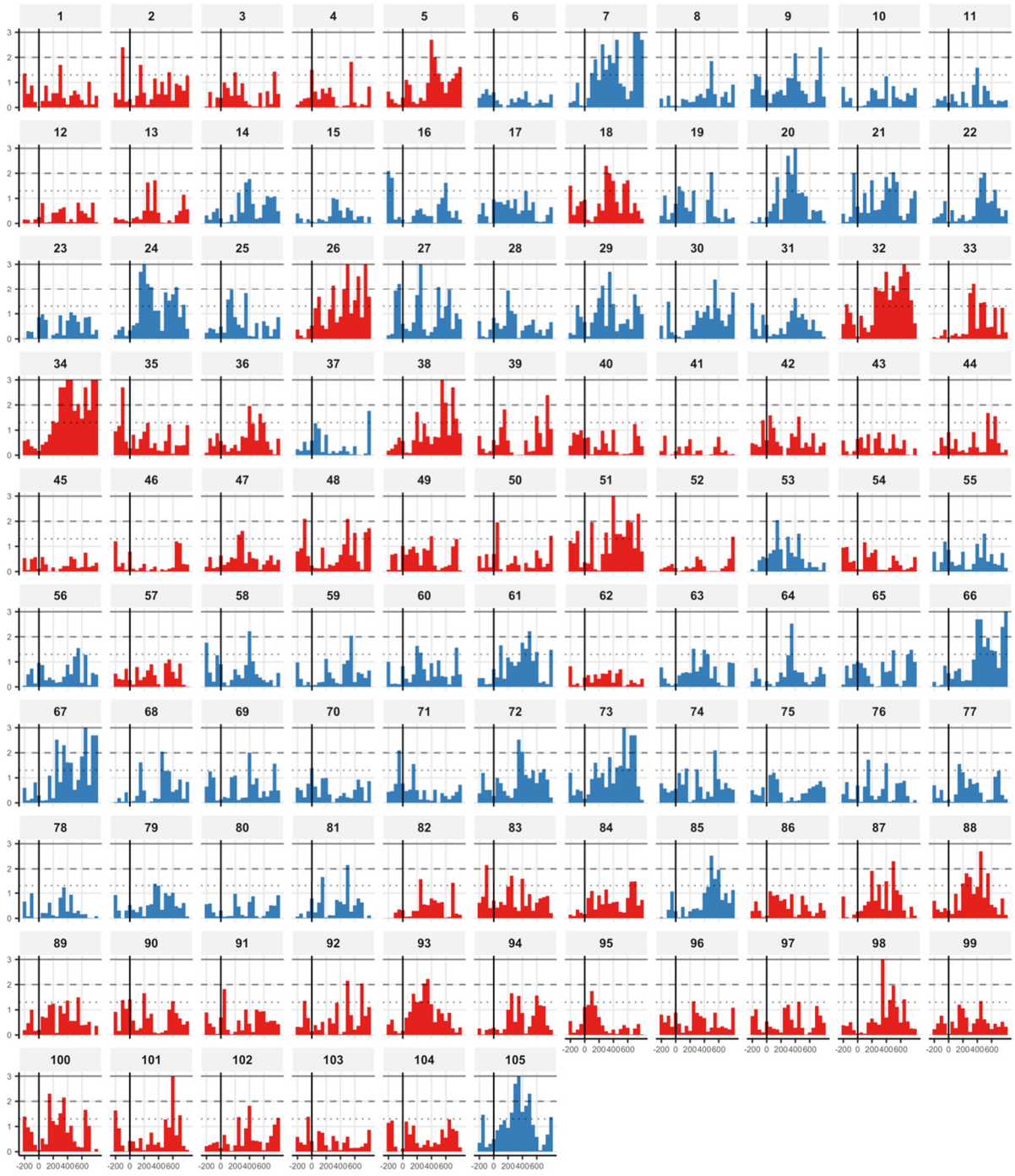
Supplementary Figure 2B. Participant-wise $-\log_{10} p$ value time series of the time-frequency-resolved EEG decoding in the theta frequency band.

The statistical significance of within-participant EEG decoding per time bin was assessed using 1,000 permutations. The x-axis represents time (in ms) relative to stimulus onset; the y-axis represents $-\log_{10} p$ values. The solid black vertical line denotes stimulus onset. The dotted horizontal line at y ≈ 1.3 denotes $p = 0.05$, the dashed horizontal line at y = 2 denotes $p = 0.01$, and the solid horizontal line at y = 3 denotes $p = 0.001$. Absolute pitch musicians are colored in red; relative pitch musicians are colored in blue.



Supplementary Figure 3A. Participant-wise decoding time series of the time-frequency-resolved EEG decoding in the beta frequency band.

The x-axis represents time (in ms) relative to stimulus onset; the y-axis represents decoding accuracy (in %). The solid black vertical line denotes stimulus onset. The solid black horizontal line at y = 50 denotes chance level (50%). Absolute pitch musicians are colored in red; relative pitch musicians are colored in blue.



Supplementary Figure 3B. Participant-wise $-\log_{10} p$ value time series of the time-frequency-resolved EEG decoding in the beta frequency band.

The statistical significance of within-participant EEG decoding per time bin was assessed using 1,000 permutations. The x-axis represents time (in ms) relative to stimulus onset; the y-axis represents $-\log_{10} p$ values. The solid black vertical line denotes stimulus onset. The dotted horizontal line at y ≈ 1.3 denotes $p = 0.05$, the dashed horizontal line at y = 2 denotes $p = 0.01$, and the solid horizontal line at y = 3 denotes $p = 0.001$. Absolute pitch musicians are colored in red; relative pitch musicians are colored in blue.

Chapter 4 Empirical Work – Article 3

Absolute and relative pitch processing in the human brain: neural and behavioral evidence

Simon Leipold, Christian Brauchli, Marielle Greber, & Lutz Jäncke

4.1 Abstract

Pitch is a primary perceptual dimension of sounds and is crucial in music and speech perception. When listening to melodies, most humans encode the relations between pitches into memory using an ability called relative pitch (RP). A small subpopulation, almost exclusively musicians, preferentially encode pitches using absolute pitch (AP): the ability to identify the pitch of a sound without an external reference. In this study, we recruited a large sample of musicians with AP (*AP musicians*) and without AP (*RP musicians*). The participants performed a pitch-processing task with a Listening and a Labeling condition during functional magnetic resonance imaging. General linear model analysis revealed that while labeling tones, AP musicians showed lower blood oxygenation level dependent (BOLD) signal in the inferior frontal gyrus and the presupplementary motor area — brain regions associated with working memory, language functions, and auditory imagery. At the same time, AP musicians labeled tones more accurately suggesting that AP might be an example of neural efficiency. In addition, using multivariate pattern analysis, we found that BOLD signal patterns in the inferior frontal gyrus and the presupplementary motor area differentiated between the groups. These clusters were similar, but not identical compared to the general linear model-based clusters. Therefore, information about AP and RP might be present on different spatial scales. While listening to tones, AP musicians showed increased BOLD signal in the right planum temporale which may reflect the matching of pitch information with internal templates and corroborates the importance of the planum temporale in AP processing. Taken together, AP and RP musicians show diverging frontal activations during Labeling and, more subtly, differences in right auditory activation during Listening. The results of this study do not support the previously reported importance of the dorsolateral prefrontal cortex in associating a pitch with its label.

This article was originally published in:

Leipold, S., Brauchli, C., Greber, M., & Jäncke, L. (2019). Absolute and relative pitch processing in the human brain: neural and behavioral evidence. *Brain Structure and Function*, *224*(5), 1723-1738. https://doi.org/10.1007/s00429-019-01872-2

4.2 Introduction

Pitch is a primary perceptual dimension of sounds and plays a crucial role in music and speech perception (Plack et al., 2005). In humans, there exist differential mechanisms to encode pitches into memory. Most individuals encode pitches in relation to other pitches using an ability called relative pitch (RP). With the exception of individuals suffering from amusia (tone deafness), all humans are able to identify changes in pitch contour by making higher-lower judgements — even from a very young age (Plantinga & Trainor, 2005). Trained musicians can also identify the exact musical interval (e.g., a perfect fifth) between pitches (McDermott & Oxenham, 2008). A small subpopulation, almost exclusively comprised of musicians, preferentially encodes pitches in absolute terms (Miyazaki & Rakowski, 2002). These musicians possess absolute pitch (AP), the ability to identify the pitch of a sound without an external reference (Deutsch, 2013; Levitin & Rogers, 2005; Zatorre, 2003). In the following, musicians with AP are referred to as *AP musicians* and musicians without AP as *RP musicians*.

A cognitive theory of AP, the two-component model, postulates that AP consists of two separate processes: The first component (*pitch memory*) comprises long-term representations of pitches which presumably exist in all humans to some extent. The second component (*pitch labeling*) comprises the associations between the long-term pitch representations and meaningful labels (e.g., C#). These associations exist exclusively in AP musicians (Levitin, 1994).

Although there has been a recent increase in neuroscientific AP research, the neural mechanisms underlying AP have been only partly identified. More than 20 years ago, it was first reported that AP musicians have a more pronounced left-right asymmetry of the planum temporale, a brain region located immediately posterior to Heschl's gyrus on the superior temporal plane (Schlaug et al., 1995). Follow-up studies found that this asymmetry might be driven by a smaller size of the right planum temporale in AP musicians rather than by a larger left planum temporale (Keenan et al., 2001; Wengenroth et al., 2014; Wilson, Lusher, Wan, Dudgeon, & Reutens, 2009). With regard to the neurophysiology of AP, a seminal study used positron emission tomography (PET) to investigate pitch processing in AP and RP musicians (Zatorre et al., 1998). While listening to tones, AP musicians showed a unique increase in cerebral blood flow (CBF) in the left posterior dorsolateral prefrontal cortex (DLPFC). Because this region has been implicated in associative learning (Petrides, Alivisatos, Evans, & Meyer, 1993), it was proposed that the CBF increase reflects the retrieval of the association between the pitch and its label from long-term memory. While labeling musical intervals, CBF increases in the posterior DLPFC were observed in both AP and RP musicians, but only RP musicians

showed increases in the right inferior frontal gyrus (IFG). These increases were interpreted as reflecting working memory demands related to the RP ability (Zatorre et al., 1998).

In the general population, the prevalence of AP is roughly estimated to be less than one in 10,000 (Bachem, 1955). Therefore, it is unsurprising that previous neuroscientific studies examining AP used small sample sizes. However, small samples result in low statistical power, which increases both the occurrence of false-negative and false-positive results (Button et al., 2013). As a consequence, previous neuroscientific AP studies reported inconsistent or even conflicting results. In this study, we aimed to counteract the statistical problems associated with small sample sizes by collecting and analyzing data from a large sample of musicians (n = 101). Using fMRI, we revisited the topic of pitch processing in AP and RP musicians. Similar to the aforementioned PET study, we employed a pitch-processing task comprising two experimental conditions (*Listening* vs. *Labeling*). Both AP and RP processing represented adequate strategies to solve the task due to its low difficulty (Itoh et al., 2005). Because individuals possessing AP preferentially encode pitches absolutely and non-possessors preferentially encode pitches relatively (Miyazaki & Rakowski, 2002), the task allowed us to contrast AP and RP processing by comparing AP musicians with RP musicians.

According to the two-component model, AP musicians differ from RP musicians by having an association between the long-term representation of a pitch and its label (Levitin, 1994). The retrieval of this pitch-label association might already occur during Listening and, to successfully perform the task, it must occur during Labeling (Zatorre et al., 1998). At the same time, AP musicians need not rely on working memory processes during Labeling (Itoh et al., 2005). For these reasons, we predicted smaller differences in AP musicians between Listening and Labeling both in BOLD signal responses and behavior. Because of their suggested role in AP processing, we expected an involvement of the posterior DLPFC and/or the planum temporale in AP musicians during Listening. Furthermore, we expected an involvement of the IFG in RP musicians during Labeling because of its association with working memory. Apart from conventional general linear model (GLM) analysis, we applied multivariate pattern analysis (MVPA) to the unsmoothed fMRI data to localize brain regions differentiating between AP and RP musicians. As a complement to GLM analysis, MVPA is sensitive to group-specific information being present in fine-grained voxel patterns which is not detectable using conventional analyses (Kriegeskorte & Bandettini, 2007). Additionally, and independently from the other analyses, we investigated ROIs previously associated with AP for group differences which are homogeneous across a brain region but too subtle to be detected by voxel-wise analysis. ROI analysis provides more statistical power than the voxel-wise analyses due

to the lower number of tests and thus, a less conservative correction for multiple comparisons (Poldrack, 2007).

4.3 Materials and Methods

4.3.1 Participants

Fifty-two AP musicians and 50 RP musicians completed the pitch-processing task. Due to a technical error during the fMRI data export, one participant of the AP group was excluded, leaving the data of 101 participants for data analysis. The two groups were matched for sex, handedness, age, musical experience, and intelligence (see Table 1).

Group assignment of the participants was based on self-report and confirmed by a tone-naming test (see below). Using both the information from self-report and a tone-naming test is advantageous because the assignment does not rely on an arbitrary cut-off concerning the tone-naming scores. In the rare case that a (potential) participant had indicated to be an AP musician in the initial online application form but then showed tone-naming scores around chance level (8.3%), we did not invite this participant for the imaging experiments in the laboratory. On the other hand, we did invite participants who had indicated to be RP musicians and then showed a high level of proficiency in tone-naming that was above chance level (and reiterated in the laboratory that they do not possess AP); please note that we did not regroup these participants as AP musicians. Furthermore, we statistically assessed if the group of RP musicians as a whole, and each RP musician individually, performed above chance level in the tone-naming test. On the group level, we found strong evidence that RP musicians performed better than chance (one sample t-test against 8.3%; $t(49) = 5.74$, $p < 10\text{-}6$, Cohen's $d = 0.81$). On the individual level, 56 % of the RP musicians performed above chance level according to a binomial test for each individual participant. Figure 1A shows the distribution of tone-naming scores. It is plausible that RP musicians performing above chance level used an internal reference (e.g., tuning standard 440 Hz) in combination with RP processing (or another yet unknown strategy) to solve the tone-naming test. Within RP musicians, tone naming did not correlate with age of onset of musical training (Pearson's $r = 0.06$, $p = 0.67$) or with cumulative musical training ($r = 0.17$, $p = 0.22$).

Table 1. Participant characteristics.

Continuous measures given as mean ± standard deviation.

	AP musicians	RP musicians
Number of participants	51	50
Sex (female / male)	23 / 28	24 / 26
Handedness (right / left / both)	45 / 4 / 2	45 / 4 / 1
Age	26.22 ± 4.91 years	25.22 ± 4.43 years
Tone-naming score	76.76 ± 20.00 %	23.93 ± 19.26 %
Musical aptitude (AMMA) - total	65.94 ± 6.20	63.32 ± 6.97
Musical aptitude (AMMA) - tonal	32.27 ± 3.68	30.48 ± 4.24
Musical aptitude (AMMA) - rhythm	33.67 ± 2.79	32.84 ± 3.03
Age of onset of musical training	6.12 ± 2.38	6.52 ± 2.42
Cumulative musical training	16111.52 ± 12590.62 hours	13903.90 ± 10072.36 hours
Crystallized intelligence (MWT-B)	27.62 ± 5.23	29.10 ± 4.72
Fluid intelligence (KAI)	124.03 ± 32.06	134.48 ± 26.91

Abbreviations: AMMA = Advanced Measures of Music Audiation, AP = absolute pitch, KAI = Kurztest für allgemeine Basisgrößen der Informationsverarbeitung, MWT-B = Mehrfachwahl-Wortschatz-Intelligenztest, RP = relative pitch.

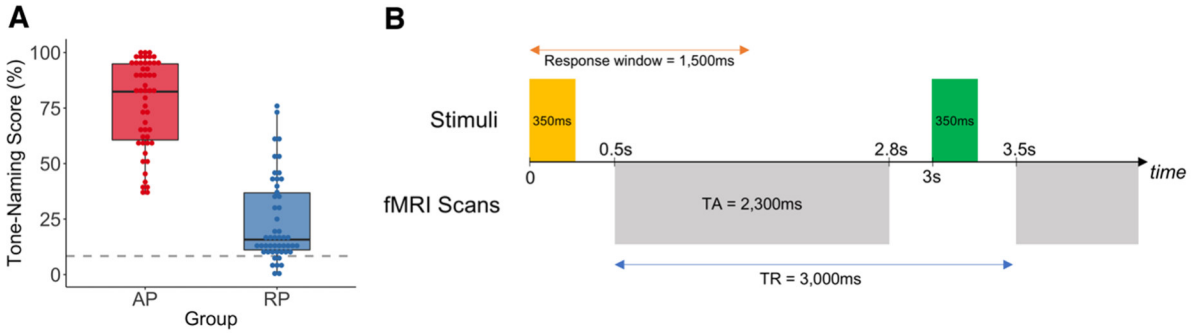


Figure 1. Tone-naming proficiency and fMRI task design.

(*A*) Distribution of group-wise tone-naming scores. The dashed grey line represents chance level (8.3%). (*B*) A single trial of the fMRI task consisted of stimulus presentation, response, and scan acquisition. The TR was longer than the TA of a single scan, so stimuli could be presented in silence. AP = absolute pitch, RP = relative pitch, TA = acquisition time, TR = repetition time.

All participants were either music professionals, music students, or highly trained amateurs between 18 and 37 years. Participants were recruited in the context of a larger project investigating AP (Brauchli et al., 2019; Burkhard et al., 2019; Greber et al., 2018; Leipold, Oderbolz, et al., 2019), which involved multiple experiments using different imaging modalities (MRI, EEG). None of the participants reported any neurological, audiological, or severe psychiatric disorders, substance abuse, or other contraindications for MRI. The absence of hearing loss was confirmed by pure tone audiometry (ST20, MAICO Diagnostics GmbH,

Berlin, Germany). Demographical data (sex, age, handedness) and part of the behavioral data (tone-naming proficiency, musical aptitude, and musical experience) was collected with an online survey tool (www.limesurvey.org). Self-reported handedness was confirmed using a German translation of the Annett questionnaire (Annett, 1970). Musical aptitude was measured using the Advanced Measures of Music Audiation (AMMA) (Gordon, 1989). Crystallized intelligence was estimated in the laboratory using the Mehrfachwahl-Wortschatz-Intelligenztest (MWT-B) (Lehrl, 2005) and fluid intelligence was estimated using the Kurztest für allgemeine Basisgrößen der Informationsverarbeitung (KAI) (Lehrl et al., 1991). All participants provided written informed consent and were paid for their participation. The study was approved by the local ethics committee (www.kek.zh.ch) and conducted according to the principles defined in the Declaration of Helsinki.

4.3.2 Sample Size Determination

We did not conduct a formal power analysis to determine the sample size for a given effect size and given power in advance of data acquisition. As data from AP musicians is extremely difficult to acquire due to their rarity (see Introduction), it was not possible to realistically plan for a specific number of participants to recruit. We rather recruited as many AP musicians as possible within a period of two years, given the limited financial and human resources available. The number of RP musicians was continuously updated to match the number of AP musicians already recruited at that time. With our final sample of about 50 participants per group, we had > 80% power to detect moderate to large effects (Cohen's $d > 0.6$) in a two sample t-test setting. Please note that power analyses in the context of neuroimaging studies are difficult to perform (Mumford, 2012), as effect sizes from previous smaller studies are probably inflated (Ioannidis, 2008; Poldrack et al., 2017), and power analyses based on pilot studies are biased (Albers & Lakens, 2018).

4.3.3 Tone-Naming Test

Participants completed a tone-naming test to assess their tone-naming proficiency (Oechslin et al., 2010). The test was carried out online at home and participants were instructed to do the test in a silent environment where they could not be disturbed. During the test, 108 pure tones were presented in a pseudorandomized order. Each tone from C3 to B5 (twelve-tone equal temperament tuning, A4 = 440 Hz) was presented three times. The tones had a duration of 500 ms and were masked with Brownian noise (duration = 2000 ms), which was presented immediately before and after the tone. Participants were instructed to identify both the chroma and the octave of the tones (e.g., C4) within 15 s of tone presentation. To calculate a score of

tone-naming proficiency, the percentage of correct chroma identifications was used. Octave errors were disregarded (Deutsch, 2013). Therefore, the chance level identification performance was at 8.3%.

4.3.4 Experimental Procedure

During fMRI scanning, participants performed a pitch-processing task (Itoh et al., 2005; Zatorre et al., 1998). The auditory stimuli used in the task consisted of three pure tones with different frequencies, and a segment of pink noise. The frequencies of the pure tones were 262 Hz (C4 in twelve-tone equal temperament tuning), 294 Hz (D4), and 330 Hz (E4). The pure tones and the noise segment had a duration of 350 ms with a 10 ms linear fade-in and a 50 ms linear fade-out. Therefore, all stimuli had an identical temporal envelope. The stimuli were created using Audacity (version 2.1.2, www.audacityteam.org). The pure tones and noise segments were presented via MRI-compatible headphones (NordicNeuroLab AS, Bergen, Norway).

The fMRI task was constructed as a rapid event-related design: Stimuli were presented in a randomized order and empty trials (without an auditory stimulus) were used to increase the efficiency of the design (Henson, 2007). Within a trial, first, a stimulus (pure tone or noise segment) was presented for 350 ms; the participants were given 1,500 ms from stimulus onset to respond. Then, 500 ms after stimulus onset, we acquired a functional scan for 2,300 ms. Finally, the trial ended with 200 ms silence before the next trial began. Due to the prolonged repetition time (TR) of 3,000 ms between two scans in comparison with the acquisition time (TA) of 2,300 ms, the stimuli were presented in the silent period (700 ms) between the acquisitions of two subsequent scans. Therefore, there was no interference of scanner noise on the perception of the stimuli (Eden, Joseph, Brown, Brown, & Zeffiro, 1999; Shah et al., 2000). The inter-trial interval (between two auditory stimuli) was varied using a jitter consisting of multiples of the TR (1–4 TRs). A visualization of the fMRI task is given in Figure 1B.

There were four runs in total. In each run, 39 pure tones (13 per chroma) and 39 noise segments were presented. The order of the stimuli was kept constant across the runs. Therefore, the auditory stimulation was identical in all runs. During the whole task, a black fixation cross on a gray background was presented on a screen. Stimulus presentation was controlled by Presentation software (version 17.1, www.neurobs.com). All stimuli and the stimulus presentation scripts are available online on the Open Science Framework (https://osf.io/ybghd/).

The task consisted of two experimental conditions: a Listening condition and a Labeling condition. These conditions only differed in the instructions given to the participants. In the

Listening condition, participants had to press one response pad button (right middle finger) when they heard a pure tone, and another button (right index finger) when they had heard a noise segment. In the Labeling condition, participants had to label the pure tones by pressing one of three corresponding buttons on the response pad (right middle, ring, and little finger in response to C4, D4, and E4, respectively) and another button (right index finger) when they had heard a noise segment. The participants were instructed not to verbally respond and to respond as quickly and as accurately as possible. The accuracy of the responses and the response time were recorded via the response pad (4 button curved right, Current Designs INC, Philadelphia, PA, USA). Both conditions lasted for two runs each. The Listening condition always preceded the Labeling condition to avoid spillover effects from the Labeling onto the Listening condition. If the order had been the other way around, AP musicians might have been tempted to still covertly label the tones in the Listening condition.

4.3.5 Statistical Analysis of Behavioral Data

In-scanner behavioral measures (response accuracy and response time) were analyzed in R (version 3.3.2, www.r-project.org). Separately for each measure, we performed a mixed-design ANOVA with a within-subject factor Condition (Listening vs. Labeling) and a between-subject factor Group (AP vs. RP). Subsequently, the two measures were separately compared within each condition using Welch's t-tests. Next, we calculated differences in both measures by subtracting the Listening from the Labeling condition for each subject. These differences were then compared between the groups again using Welch's t-tests. Finally, the differences were correlated with the tone-naming scores using the Pearson correlation coefficient. The significance level was set to $p < 0.05$. Generalized eta-squared (η^2_G) was used as an effect size for effects within an ANOVA and Cohen's *d* for t-tests.

4.3.6 Imaging Data Acquisition and Preprocessing

Imaging data was acquired on a Philips Ingenia 3.0 T MRI system (Philips Medical Systems, Best, The Netherlands), equipped with a commercial 15-channel head coil. Whole-brain functional images were acquired in four runs using a T2*-weighted gradient echo (GRE) echo planar imaging (EPI) sequence (scan duration of one run = 380 s). The T2*-weighted sequence had the following parameters: TR = 3000 ms, TA = 2300 ms, echo time (TE) = 35 ms, flip angle $\alpha = 90^{o}$, number of axial slices = 38, slice gap = 0.6 mm, slice scan order = interleaved, field of view (FOV) = 220 x 220 x 136 mm^3, acquisition voxel size = 3.0 x 3.0 x 3.0 mm^3, reconstructed voxel size = 2.75 x 2.75 x 3.6 mm^3, reconstruction matrix = 80 x 80, number of dummy scans = 3, total number of scans = 122.

In addition, a whole-brain structural image was acquired using a T1-weighted GRE turbo field echo sequence (scan duration = 350 s). The T1-weighted sequence had the following parameters: TR = 8100 ms, TE = 3.7 ms, flip angle $\alpha = 8^{\circ}$, number of sagittal slices = 160, FOV = 240 x 240 x 160 mm^3, acquisition voxel size = 1.0 x 1.0 x 1.0 mm^3, reconstructed voxel size = 0.94 x 0.94 x 1.0 mm^3, reconstruction matrix = 256 x 256. The whole scanning session lasted around 50 minutes and also involved resting-state fMRI and DTI. The results of these imaging modalities are discussed in other publications.

The functional images and the structural images were preprocessed using SPM12 (version 6906, www.fil.ion.ucl.ac.uk/spm/software/spm12). The following preprocessing steps were performed in succession using default settings unless otherwise stated: (i) Slice time correction. (ii) Motion correction by a rigid body transformation using six parameters (three translations and three rotations). We did not use unwarping as we had not collected data to correct geometrical distortions caused by susceptibility-induced magnetic field inhomogeneities. (iii) Coregistration of the structural image to the mean functional image. (iv) Segmentation and bias field correction of the structural image and estimation of the deformation field to map the image to the T1-weighted MNI152 template. (v) Normalization of the functional images using the estimated deformation field. (vi) Interpolation to an isotropic voxel size of 3.0 mm. (vii) Smoothing of the functional images with an 8 mm full width at half maximum (FWHM) three-dimensional Gaussian kernel. The quality of the normalization was visually inspected to confirm proper execution.

4.3.7 GLM Analysis

Subject-wise first-level analysis was performed in SPM12. The voxel-wise BOLD signal time series was modeled using a GLM. The first-level design matrix contained, for each run separately, two regressors of interest (onsets of pure tones, onsets of noise segments) and one regressor of no interest (onsets of button presses). These regressors were modeled by convolving delta functions with the canonical double-gamma hemodynamic response function (HRF). Furthermore, we included the six motion parameters estimated during preprocessing as nuisance regressors and applied a high-pass filter (cutoff = 128 s) to remove low-frequency drifts. The following first-level contrasts of interest were calculated: $\text{Tones}_{\text{Listening}} > \text{Noise}_{\text{Listening}}$ and $\text{Tones}_{\text{Labeling}} > \text{Noise}_{\text{Labeling}}$. Following the logic of cognitive subtraction, these contrasts reflect BOLD signal increases associated with pitch processing.

Second-level random effects analysis was performed using non-parametric permutation tests as implemented in SnPM13 (www.warwick.ac.uk/snpm). Permutation tests depend on fewer

assumptions than standard parametric approaches and provide an exact control of the family-wise error (FWE) rate (Eklund, Nichols, & Knutsson, 2016; Nichols & Holmes, 2002). For the second-level analysis, we used a 2 x 2 mixed factorial design to investigate the interaction between Group (AP vs. RP) and Condition (Listening vs. Labeling). To facilitate the interpretation of the interaction, difference images were created for each subject by subtracting the contrast image of the Listening condition (Tones $_{\text{Listening}}$ > Noise $_{\text{Listening}}$) from the contrast image of the Labeling condition (Tones $_{\text{Labeling}}$ > Noise $_{\text{Labeling}}$). These difference images were entered in SnPM13 as inputs for a two sample t-test to compare AP and RP musicians (cluster-wise inference, 10000 permutations, cluster defining threshold (CDT) $p < 0.001$). An anatomically defined mask was used to restrict the search space of the analysis to a priori defined brain regions previously associated with AP and RP processing. To create this mask, we used probability maps of the following bilateral brain regions included in the Harvard-Oxford cortical atlas (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases). (i) Heschl's gyrus, (ii) planum temporale, (iii) planum polare, (iv) superior temporal gyrus (anterior and posterior division), (v) superior frontal gyrus, (vi) middle frontal gyrus, (vi) inferior frontal gyrus (pars opercularis and pars triangularis), (vii) superior parietal lobule, (ix) gyrus supramarginalis (anterior and posterior division), and (x) angular gyrus. The probability maps were then combined, thresholded and binarized at 10% probability using the utility fslmaths (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Fslutils). Using a mask to restrict the search space alleviates the problem of multiple comparisons as less voxels are tested for an effect (Poldrack, 2007). This particular mask furthermore reflects prior knowledge that has been accumulated about AP and RP in many studies over the years. Structural and functional alterations of auditory regions in the superior temporal cortex have been repeatedly linked to AP processing (Brauchli et al., 2019; Jäncke, Langer, & Hänggi, 2012; Keenan et al., 2001; Kim & Knösche, 2016, 2017; McKetton, DeSimone, & Schneider, 2019; Schlaug et al., 1995; Schulze et al., 2013; Wengenroth et al., 2014; Wilson et al., 2009). Dorsal and ventral frontal areas have been associated with both AP and RP (Bermudez, Lerch, Evans, & Zatorre, 2009; Brauchli et al., 2019; Dohn et al., 2015; Ohnishi et al., 2001; Wengenroth et al., 2014; Zatorre et al., 1998), and there is evidence that parietal areas contribute to AP (Brauchli et al., 2019; Loui et al., 2012) and RP (Foster & Zatorre, 2010a, 2010b; Schulze et al., 2009).

Two follow-up analyses with the same mask were performed. To determine the effects of condition within each group, we entered the difference images as inputs for a one sample t-test for each group separately (cluster-wise inference, 10000 permutations, CDT $p < 0.001$). To determine the effects of group within each condition, we entered the first-level contrast images

(Tones $_{Listening}$ > Noise $_{Listening}$, Tones $_{Labeling}$ > Noise $_{Labeling}$) as inputs for a one sample t-test for each condition separately (cluster-wise inference, 10000 permutations, CDT $p < 0.001$). The significance level for all analyses was set to $\alpha = 0.05$, FWE-corrected for multiple comparisons.

Additionally, we performed a GLM-based whole-brain analysis to explore effects located outside of brain regions previously associated with AP or RP. This exploratory analysis extended the search space to all brain regions of the Harvard-Oxford cortical and subcortical atlases (excluding the cerebral white matter, the brain stem, and the lateral ventricles). In this whole-brain analysis, we employed the same second-level analysis steps as described above for the restricted analysis.

4.3.8 MVPA

We carried out a specific type of MVPA, namely searchlight analysis as implemented in PyMVPA (version 2.6.1, www.pymvpa.org) to detect brain regions containing fine-grained BOLD signal patterns which differentiated between AP and RP musicians (Etzel, Zacks, & Braver, 2013; Kriegeskorte, Goebel, & Bandettini, 2006). Due to the high computational demands, all analyses were carried out on the ScienceCloud of the University of Zurich (www.s3it.uzh.ch). Searchlight analysis, sometimes called information-based brain mapping, builds a map of voxels which are informative regarding group status (searchlight analysis can also be used to analyze information about different stimuli or experimental conditions). A machine learning classifier uses local BOLD signal patterns to classify the participants as belonging to one of the two groups. Brain regions which contain clusters of informative voxels are differentially activated in the two groups (Kriegeskorte & Bandettini, 2007; Kriegeskorte et al., 2006).

Searchlight analysis was performed on the unsmoothed functional images. To some extent, smoothing removes the fine-grained patterns of activation which were here analyzed for information about group status (Kriegeskorte & Bandettini, 2007). Analogous to the GLM analysis, two first-level contrasts were computed in SPM12 (this time using the unsmoothed images): Tones $_{Listening}$ > Noise $_{Listening}$ and Tones $_{Labeling}$ > Noise $_{Labeling}$. In addition, we again calculated a difference image for each subject by subtracting the contrast image of the Listening condition from the contrast image of the Labeling condition.

In total, we performed three searchlight analyses using the different images (difference images, Listening contrast images, Labeling contrast images) as inputs. In all analyses, a sphere was moved across all voxels of the anatomically defined mask that was also used in the GLM analysis. Each sphere had a radius of three voxels (9 mm) and consisted of one center voxel

and (at most) 122 surrounding voxels. In every sphere, a linear support vector machine (C = 1) was trained and tested using a 5-fold cross-validation. For the cross-validation, the input images were pseudorandomly partitioned into five chunks under the restriction that each chunk contained the same number of images of AP musicians and RP musicians. One chunk contained 11 images of AP musicians (instead of 10), because our analyzed sample included 51 AP and 50 RP musicians. The average classification accuracy of the five folds was written in the location of the center voxel to create a map of classification accuracies (i.e. an information map).

To assess the statistical significance of informative clusters, we used non-parametric permutation testing (Nichols & Holmes, 2002). For this purpose, each of the three searchlight analyses was repeated with permuted group labels (10000 permutations). For every iteration, the group labels were randomly permuted within each chunk. We used this restriction to balance the number of images per group in each chunk. The resulting permutation set was fixed for the whole searchlight analysis (i.e. across all center voxels of the mask) to preserve the spatial dependency between neighboring center voxels (Stelzer, Chen, & Turner, 2013). All properties of the searchlight analyses with the permutated labels were identical to the analyses with the real labels (e.g., classifier parameters, cross-validation scheme). The permutation procedure resulted in a null distribution of 10000 information maps.

Next, both the empirical information map (created with the real labels) and the null information maps (created with the permuted labels) were thresholded with a CDT of $p < 0.001$ using custom MATLAB R2016a functions. Subsequently, we formed clusters of the above-threshold voxels using CoSMoMVPA (version 1.1.0, www.cosmomvpa.org). The maximum cluster size of each null information map was extracted to form a null distribution of cluster sizes. Finally, the p value of the clusters in the empirical information map was calculated as the proportion of cluster sizes under the null distribution that were larger than the empirical cluster size. The significance level was set to $\alpha = 0.05$, FWE-corrected.

4.3.9 ROI Analysis

In addition to the voxel-wise GLM and searchlight analyses, the mean BOLD signal changes in a priori defined ROIs were compared between groups using MarsBaR (version 0.44, www.marsbar.sourceforge.net). We defined four ROIs which have been previously associated with AP processing: left planum temporale (Schlaug et al., 1995; Wilson et al., 2009), right planum temporale (Keenan et al., 2001; Wengenroth et al., 2014; Wilson et al., 2009), left

DLPFC (Bermudez & Zatorre, 2005; Ohnishi et al., 2001; Zatorre et al., 1998), and right DLPFC (Bermudez & Zatorre, 2005).

The ROIs were created as spheres (radius = 10 mm) based on MNI coordinates. We used anatomically defined coordinates for the planum temporale and functionally defined coordinates for the DLPFC, because the planum temporale can be delineated by anatomical landmarks, whereas the DLPFC is primarily a functional region. The coordinates of the left (x = -44, y = -34, z = 11) and right planum temporale (x = 41, y = -31, z = 15) were derived from the Harvard-Oxford cortical atlas planum temporale probability map by choosing the voxel with the highest probability in the left and the right hemisphere. The coordinates of the left DLPFC (x = -40, y = 9, z = 42) were taken from a seminal study investigating pitch processing in AP, which was the first to associate this brain region with the retrieval of the pitch-label association while AP musicians were listening to tones (Zatorre et al., 1998). The original study reported the coordinates in Talairach space, so we transformed the coordinates into MNI space (Lacadie, Fulbright, Rajeevan, Constable, & Papademetris, 2008). The coordinates of the left hemispheric region were flipped at the midsagittal plane to derive the coordinates of the right DLPFC (x = 40, y = 9, z = 42). For each subject and ROI, we extracted first-level contrast values from the Listening condition (Tones Listening > Noise Listening). For each ROI, these contrast values were compared between AP and RP musicians using Welch's t-tests in R. The significance level was set to $\alpha = 0.0125$, FWE-corrected for multiple ROIs.

At the request of a reviewer, we conducted a supplemental ROI analysis for the bilateral Heschl's gyrus. Recent studies have implicated both structure and function of Heschl's gyrus in AP processing (Brauchli et al., 2019; McKetton et al., 2019; Wengenroth et al., 2014). Coordinates for left (x = -44, y = -24, z = 11) and right Heschl's gyrus (x = 42, y = -20, z = 9) were anatomically defined, analogous to the planum temporale, by choosing the voxel with the highest probability per hemisphere of the Harvard-Oxford cortical atlas Heschl's gyrus probability map. In line with the exploratory character of this analysis, here, we used a significance level of $\alpha = 0.05$, uncorrected.

4.4 Results

4.4.1 Behavior

Demographical and behavioral characteristics of the AP musicians (n = 51) and the RP musicians (n = 50) were compared using Welch's t-tests. The two groups did not differ in age ($t(98.3) = 1.07$, $p = 0.29$), age of onset of musical training ($t(98.9) = -0.84$, $p = 0.40$), cumulative musical training ($t(95.19) = 0.97$, $p = 0.33$), crystallized intelligence ($t(96.4) = -1.48$, $p = 0.14$),

and fluid intelligence (t(96.7) = -1.78, $p = 0.08$). As predicted, AP musicians had a substantially higher tone-naming score than RP musicians (t(99) = 13.53, $p < 10^{-15}$). There was a trend towards a higher musical aptitude in AP musicians as quantified by the AMMA total score (t(97.2) = 1.99, $p = 0.05$). Follow-up analyses of the AMMA subscores showed that this difference was driven by a slightly higher tonal score in AP musicians (t(96.5) = 2.27, $p = 0.03$), but there was no difference regarding the rhythm score (t(98.0) = 1.42, $p = 0.16$). Descriptive statistics of participant characteristics are given in Table 1.

The in-scanner behavioral measures were analyzed using a mixed-design ANOVA with a within-subject factor Condition (Listening vs. Labeling) and a between-subject factor Group (AP vs. RP). As shown in Figure 2A, the mixed-design ANOVA of the response accuracy revealed an interaction between the factors Group and Condition ($F(1,99) = 8.37$, $p = 0.005$, $\eta^2_G = 0.02$). The difference in response accuracy between the two conditions (Labeling minus Listening) was smaller in AP than in RP musicians (Welch's t-test, t(79.1) = 2.88, $p = 0.005$, $d = 0.57$). Furthermore, this difference correlated with the tone-naming score ($r = 0.41$, $p < 0.001$). On average, the response accuracy was higher in the Listening condition than in the Labeling condition, so this correlation indicates a smaller difference for participants with a higher tone-naming score (see Figure 2C). Additional follow-up analyses showed a higher response accuracy for AP musicians in the Labeling condition (Welch's t-test, t(73.4) = 2.88, $p = 0.005$, $d = 0.57$), but not in the Listening condition (Welch's t-test, t(87.7) = 1.10, $p = 0.28$, $d = 0.22$). As shown in Figure 2B, the mixed-design ANOVA of the response time revealed a Group x Condition interaction ($F(1,99) = 8.85$, $p = 0.004$, $\eta^2_G = 0.01$). The condition difference in response time was smaller in AP musicians (Welch's t-test, t(95.6) = -2.97, $p = 0.004$, $d = 0.59$). Again, this difference correlated with the tone-naming score ($r = -0.31$, $p = 0.002$) (see Figure 2D). Descriptive statistics of the in-scanner behavioral measures are given in Table 2.

Table 2. In-scanner behavioral measures.

Measures given as mean ± standard deviation.

	AP musicians	**RP musicians**
Response accuracy Listening	96.69 ± 2.68 %	95.97 ± 3.82 %
Response accuracy Labeling	96.88 ± 3.14 %	94.12 ± 6.04 %
Response time Listening	550.23 ± 125.60 ms	516.20 ± 114.61 ms
Response time Labeling	642.33 ± 145.40 ms	649.29 ± 139.13 ms

Abbreviations: AP = absolute pitch, RP = relative pitch

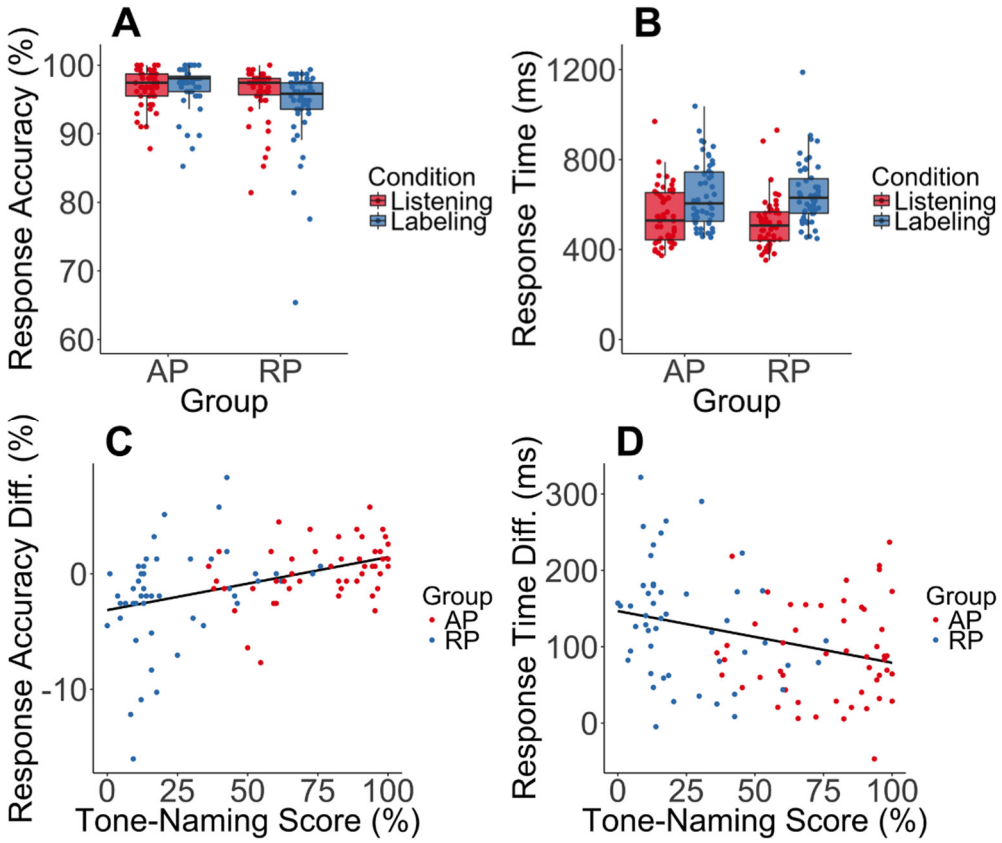


Figure 2. In-scanner behavioral measures (response accuracy and response time).

(A) Interaction between Group (AP vs. RP) and Condition (Listening vs. Labeling) as revealed by a mixed-design ANOVA of the response accuracy ($F(1,99) = 8.37$, $p = 0.005$, $\eta^2_G = 0.02$). The interaction is characterized by smaller differences between Listening and Labeling in AP musicians than RP musicians ($t(79.1) = 2.88$, $p = 0.005$, $d = 0.57$). Additionally, AP musicians demonstrated higher response accuracy in the Labeling condition ($t(73.4) = 2.88$, $p = 0.005$, $d = 0.57$). (B) Group x Condition interaction as revealed by a mixed-design ANOVA of response time ($F(1,99) = 8.85$, $p = 0.004$, $\eta^2_G = 0.01$), again characterized by smaller condition differences in AP musicians ($t(95.6) = -2.97$, $p = 0.004$, $d = 0.59$). (C) Correlation between the condition difference in response accuracy (Labeling minus Listening) and tone-naming score ($r = 0.41$, $p < 0.001$). Note that the positive correlation indicates a smaller difference for participants with a higher tone-naming score. (D) Correlation between the condition difference in response time and tone-naming score ($r = -0.31$, $p = 0.002$). AP = absolute pitch, RP = relative pitch.

4.4.2 BOLD Signal Changes

The BOLD signal changes were analyzed using a voxel-wise GLM in combination with a second-level mixed factorial design., Using the mask restricting the search space to brain regions previously associated with AP or RP, we found a Group x Condition interaction which was characterized by smaller BOLD signal condition differences in AP musicians, paralleling the in-scanner behavioral measures. As shown in Figure 3A, this interaction was detected in three frontal clusters (see Table 3 for details). FWE-corrected p values (p_{FWE}) and the number of voxels (k) of clusters are given in brackets. The clusters were localized in the right IFG, pars opercularis ($p_{FWE} < 0.001$, $k = 407$) and the left IFG, pars opercularis ($p_{FWE} = 0.003$, $k = 169$). A third cluster was localized in the presupplementary motor area (preSMA) of the dorsomedial

prefrontal cortex (p_{FWE} = 0.005, k = 141). The exploratory whole-brain analysis for the Group x Condition interaction yielded virtually identical clusters with the same maxima. These clusters were slightly more extended than in the restricted analysis. For full transparency, we made the unthresholded *t*-maps of the whole-brain analyses available on NeuroVault (Gorgolewski et al., 2015), https://neurovault.org/collections/4906/.

Table 3. Group x Condition interaction of BOLD signal

The coordinates (x, y, z) are in MNI space. The clusters are ordered according to their size.

Contrast	Brain Region	k	t_{max}	x	y	z	P_{FWE}
AP < RP	Right IFG, pars opercularis	407	5.90	48	8	14	< 0.001
AP < RP	Left IFG, Pars opercularis	169	5.00	-54	11	5	0.003
AP < RP	preSMA	141	4.23	-6	17	47	0.005

Abbreviations: AP = absolute pitch, IFG = inferior frontal gyrus, k = number of voxels, preSMA = presupplementary motor area, RP = relative pitch

As shown in Figure 3B and 3C, using the restricted search space, follow-up analyses within each group separately revealed similar BOLD signal differences between the two conditions with the exception of the three clusters described above (bilateral IFG, preSMA). In the bilateral IFG and the preSMA, only RP musicians showed increased BOLD signal in the Labeling condition. In addition, both groups showed increases in the bilateral intraparietal sulcus (IPS) and the bilateral DLPFC (see Table 4). These increases were stronger and more distributed in RP musicians, again indicating larger condition differences. Further follow-up analyses within each condition revealed that there were no group differences in the Listening condition (see Figure 3D). In contrast, AP musicians showed lower BOLD signal in the Labeling condition in the right IFG (p_{FWE} < 0.001, k = 312), the left IFG (p_{FWE} = 0.003, k = 195), and the preSMA (p_{FWE} = 0.005, k = 134). These clusters were equivalent to the clusters of the Group x Condition interaction (see Figure 3E and Table 5). The whole-brain analysis yielded again virtually identical results with slightly extended clusters. Unthresholded *t*-maps of the whole-brain follow-up analyses are available on NeuroVault (https://neurovault.org/collections/4906/).

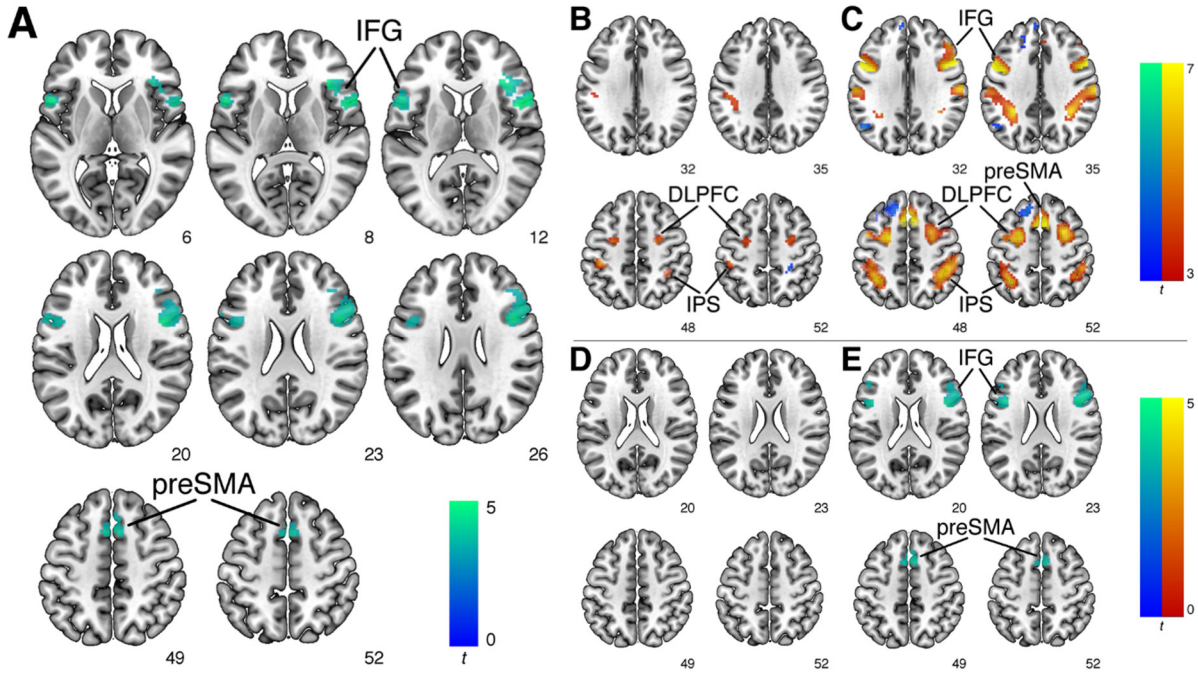


Figure 3. Results of the GLM analysis.

(*A*) Group x Condition interaction characterized by smaller condition differences in AP musicians in the right inferior frontal gyrus (IFG), pars opercularis ($p_{FWE} < 0.001$, $k = 407$), left IFG, pars opercularis ($p_{FWE} = 0.003$, $k = 169$), and presupplementary motor area (preSMA) of the dorsomedial prefrontal cortex ($p_{FWE} = 0.005$, $k = 141$). Cold colors indicate AP < RP. (*B*) Follow-up analysis within AP musicians revealed increases during Labeling in bilateral intraparietal sulcus (IPS) and bilateral dorsolateral prefrontal cortex (DLPFC). Hot colors indicate Labeling > Listening and cold colors indicate Listening > Labeling. (*C*) Follow-up analysis within RP musicians revealed similar increases during Labeling in bilateral IPS and bilateral DLPFC and unique increases in the bilateral IFG and the preSMA. Hot colors indicate Labeling > Listening and cold colors indicate Listening > Labeling. (*D*) Follow-up analysis within the Listening condition revealed no group differences. (*E*) Follow-up analysis within the Labeling condition revealed equivalent clusters to the Group x Condition interaction in the right IFG ($p_{FWE} < 0.001$, $k = 312$), the left IFG ($p_{FWE} = 0.003$, $k = 195$), and the preSMA ($p_{FWE} = 0.005$, $k = 134$). Cold colors indicate AP < RP. AP = absolute pitch, RP = relative pitch.

Table 4. Condition differences in BOLD signal within each group.

The coordinates (x, y, z) are in MNI space.

Group	Contrast	Brain Region	k	t_{max}	x	y	z	P_{FWE}
AP	Labeling > Listening	Left IPS	101	5.37	-42	-31	50	0.009
AP	Labeling > Listening	Left DLPFC	69	4.56	-27	-7	47	0.01
AP	Labeling > Listening	Right DLPFC	41	4.59	27	-4	50	0.03
AP	Labeling > Listening	Right IPS	29	4.51	36	-43	44	0.04
RP	Labeling > Listening	Left IFG, pars opercularis, preSMA, Left DLPFC	1290	9.85	-57	8	23	< 0.001
RP	Labeling > Listening	Right IFG, pars opercularis, Right DLPFC	1053	10.95	54	8	14	< 0.001
RP	Labeling > Listening	Left IPS	622	7.01	-33	-46	38	< 0.001
RP	Labeling > Listening	Right IPS	595	7.14	39	-46	44	< 0.001
AP	Listening > Labeling	Right SPL	58	4.48	30	-34	62	0.02
RP	Listening > Labeling	Left DLPFC	156	5.05	-21	38	41	0.004
RP	Listening > Labeling	Left frontal pole	91	5.93	-9	56	23	0.01
RP	Listening > Labeling	Right SPL	58	4.72	12	-46	71	0.02
RP	Listening > Labeling	Left angular gyrus	45	4.97	-51	-64	32	0.03

Abbreviations: AP = absolute pitch, DLPFC = dorsolateral prefrontal cortex, IPS = intraparietal sulcus, IFG = inferior frontal gyrus, k = number of voxels, preSMA = presupplementary motor area, RP = relative pitch, SPL = superior parietal lobule.

Table 5. Group differences in BOLD signal in the Labeling condition

The coordinates (x, y, z) are in MNI space.

Contrast	Brain Region	k	t_{max}	x	y	z	P_{FWE}
AP < RP	Right IFG, pars opercularis	312	4.63	45	11	23	< 0.001
AP < RP	Left IFG, Pars opercularis	195	4.17	-42	8	23	0.003
AP < RP	preSMA	134	4.69	9	23	44	0.005

Abbreviations: AP = absolute pitch, IFG = inferior frontal gyrus, k = number of voxels, preSMA = presupplementary motor area, RP = relative pitch

4.4.3 Group Decoding by Searchlight Analysis

In addition to the voxel-wise GLM, we used searchlight analysis to localize BOLD signal patterns which differentiate between the two groups (Kriegeskorte et al., 2006). For the main analysis, we used the difference in BOLD signal patterns between the two conditions as the input. As shown in Figure 4A, group status could be decoded in the left IFG, pars triangularis (p_{FWE} = 0.01, k = 29). The mean classification accuracy within the cluster was 72.5%. In comparison to the left IFG cluster from the GLM Group x Condition interaction, this cluster was located more anteriorly on the IFG. Follow-up analyses were performed with the patterns of each condition separately. Analogous to the GLM analysis, group status could not be decoded based on patterns in the Listening condition. In contrast, group status could be decoded based on Labeling patterns in the preSMA ($p_{FWE} < 0.001$, $k = 81$, mean classification accuracy = 70.6%). This cluster substantially overlapped with the preSMA cluster from the GLM (see Figure 4A). However, a complete overlap should not be expected, because searchlight analysis is known to cause slight distortions in the localization (Etzel et al., 2013).

4.4.4 Regional Mean BOLD Signal Changes

Finally, we extracted the mean BOLD signal changes from a priori defined ROIs. The bilateral planum temporale and the bilateral DLPFC were used as ROIs as these regions have previously been associated with AP processing (Bermudez & Zatorre, 2005; Keenan et al., 2001; Ohnishi et al., 2001; Schlaug et al., 1995; Wengenroth et al., 2014; Wilson et al., 2009; Zatorre et al., 1998). It has been proposed that AP musicians automatically retrieve the pitch-label association from long-term memory when confronted with tones (Itoh et al., 2005). Therefore, the group comparison of mean BOLD signal changes was only performed in the Listening condition (Itoh et al., 2005; Ohnishi et al., 2001; Zatorre et al., 1998). As described above, we did not find group differences during Listening with the voxel-wise GLM analysis and the searchlight analysis. However, these analyses may miss subtle effects related to the automatic retrieval because of their conservative correction for multiple comparisons (Poldrack, 2007). As shown in Figure 4B, AP musicians showed increased mean BOLD signal in the right planum temporale (Welch's t-test, $t(94.6) = 2.66$, $p = 0.01$, $d = 0.53$), but not in the left planum temporale, the left DLPFC, and the right DLPFC (all $p > 0.10$). The exploratory ROI analysis of bilateral Heschl's gyrus did not reveal group differences in mean BOLD signal in the left Heschl's gyrus (p = 0.20). Also, the mean BOLD signal in the right Heschl's gyrus did not significantly differ between the groups (p = 0.09), although there was descriptively a tendency towards higher BOLD signal in AP musicians associated with a small effect size (d = 0.34).

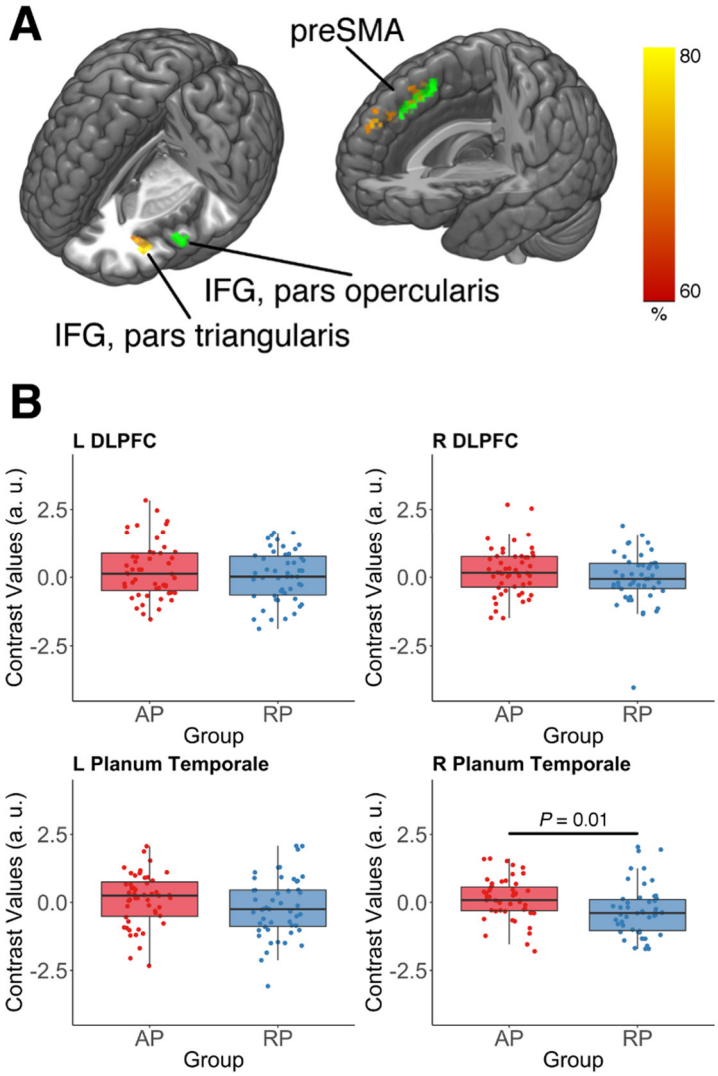


Figure 4. Results of the searchlight analysis and the ROI analysis.

(*A*) Left: Group status could be decoded in the left inferior frontal gyrus (IFG), pars triangularis ($p_{FWE} = 0.01$, $k = 29$) based on the difference in BOLD signal patterns between Listening and Labeling (shown in red-yellow). The cluster is located more anteriorly on the IFG compared to the Group x Condition cluster from the GLM analysis (shown in green). Right: Group status decoding in the presupplementary motor area (preSMA, $p_{FWE} < 0.001$, $k = 81$) based on patterns in the Labeling condition (shown in red-yellow). There is substantial overlap with the preSMA cluster revealed by the GLM group comparison during Labeling (shown in green). Hot colors represent the classification accuracy. (*B*) AP musicians show higher mean BOLD signal chances during Listening in the right planum temporale ($t(94.6) = 2.66$, $p = 0.01$, $d = 0.53$), but not in the left planum temporale or the bilateral dorsolateral prefrontal cortex (DLPFC). AP = absolute pitch, RP = relative pitch.

4.5 Discussion

In this study, we investigated AP and RP processing in the human brain using task-based fMRI in a large sample of musicians. The GLM analysis revealed smaller BOLD signal differences between Listening and Labeling in AP musicians than in RP musicians. The smaller differences between the conditions were driven by lower BOLD signals in AP musicians during Labeling in the left- and right-sided pars opercularis of the IFG and the preSMA. The in-scanner behavioral measures (response accuracy and response time) mirrored the fMRI data by showing smaller differences between Listening and Labeling in AP musicians. Using MVPA, we found that group status could be decoded in the left-sided pars triangularis of the IFG based on the difference in BOLD signal patterns between Listening and Labeling. Furthermore, group decoding was also possible in the preSMA based on BOLD signal patterns obtained in the Labeling condition. Lastly, the ROI analysis revealed a higher mean BOLD signal in AP musicians during Listening in the right planum temporale which was not detected by the GLM analysis and the MVPA.

The IFG is an important target region for auditory information which is propagated from the auditory cortex to the IFG along the ventral stream (the "what" pathway) of auditory processing (Rauschecker & Scott, 2009). In this context, the IFG has been repeatedly linked with auditory working memory functions (Schulze, Koelsch, & Williamson, 2018). More specifically, the IFG has been associated with working memory for pitch, as shown by both PET and fMRI studies (Gaab, Gaser, Zaehle, Jäncke, & Schlaug, 2003; Zatorre, Evans, & Meyer, 1994). In this study, we observed BOLD signal increases in RP musicians bilaterally in the IFG during Labeling. This increase was not observable in AP musicians. As RP musicians need to use their RP ability to successfully complete the task, it is plausible that the signal increase in the IFG reflects pitch working memory processes as an important aspect of RP processing (McDermott & Oxenham, 2008). This interpretation is fully in line with the results of the PET study described in the introduction (Zatorre et al., 1998). In this study, RP musicians, but not AP musicians, showed CBF increases in IFG while they were labeling musical intervals. More evidence for the association between RP processing and working memory comes from a number of electrophysiological studies investigating the P300 component of the auditory event-related potential. The P300 presumably reflects the updating of auditory information in working memory. Several studies found an absent or reduced P300 component in AP musicians not relying on RP processing. In contrast, RP musicians show a normal P300 amplitude (Itoh et al., 2005; Klein et al., 1984).

Apart from being implicated in working memory, the IFG has been strongly associated with language functions. In the left hemisphere, the pars opercularis (Brodmann area 44) and the pars triangularis (Brodmann area 45) of the IFG are known as Broca's area, a brain region traditionally associated with speech production, but also heavily involved in speech perception (Friederici, 2011). In the right hemisphere, the IFG is linked to the perception of prosody (pitch changes in speech) (Buchanan et al., 2000). Therefore, the BOLD signal increases in RP musicians in bilateral IFG might reflect language-related processes. More concretely, the RP musicians might have engaged in covert articulation of the tone labels as a part of their strategy to label the tones. In contrast, it seems that the AP musicians do not rely on a verbal code to successfully complete the task. This is in accordance with behavioral evidence demonstrating non-verbal coding strategies in AP musicians (Zatorre & Beckett, 1989), and fMRI evidence showing atypically similar BOLD signal in AP musicians during the perception of tonal and verbal stimulus material (Schulze et al., 2013).

Mirroring the bilateral IFG BOLD signal increases, the preSMA showed signal increases in RP musicians during Labeling. In addition, the BOLD signal patterns during Labeling in the preSMA contained information about group status. Thus, AP and RP processing were accompanied by differential BOLD signal patterns. The preSMA is anatomically connected to the IFG via the frontal aslant tract and has been implicated in speech production and processing (Catani et al., 2013). More importantly, the preSMA plays a key role in the auditory imagery of pitch (Lima, Krishnan, & Scott, 2016). Auditory imagery generally refers to the generation of auditory information in the absence of sound perception. However, auditory imagery can also involve auditory information that is generated in addition to the currently perceived information. Consequently, RP musicians might have imagined the pitches of previously heard tones to determine the pitch of the current tone. This interpretation is in line with the anecdotal observation that RP musicians often covertly sing pitches in order to identify the musical intervals. It is important to note that the working memory and the language explanations of the IFG and preSMA involvement during Labeling are not mutually exclusive. There is evidence that largely overlapping brain regions are involved in auditory working memory for verbal material and non-verbal material, for example, pitches (Koelsch et al., 2009).

The results from the GLM analysis and the MVPA did not fully converge with regard to the localization of the group differences. Most notably, using MVPA, we found that group status could be decoded from BOLD signal patterns in the left-sided pars triangularis of the IFG whereas the GLM revealed BOLD signal differences in the pars opercularis. As mentioned above, these two regions constitute Broca's area. In a previous study using MVPA, it was

shown that BOLD signal patterns in Broca's area contain speech-related information which was not detectable with GLM analysis (Lee, Turkeltaub, Granger, & Raizada, 2012). MVPA is more sensitive to information in fine-grained patterns which are preserved in unsmoothed fMRI data (Kriegeskorte & Bandettini, 2007). At the same time, there has been a debate about whether or not Broca's area should be divided into subareas executing different functions (Friederici, 2011). Consequently, we propose that the BOLD signal patterns in the pars triangularis represent information about AP and RP on a smaller spatial scale. In contrast, the differences in the pars opercularis might be more homogeneous and therefore detectable by the GLM analysis. Further studies should elucidate the potentially differential roles of these two brain regions in pitch processing.

Although showing lower BOLD signal in the IFG and preSMA during Labeling, the AP musicians identified the tones more accurately than RP musicians. Therefore, AP processing might be more efficient than RP processing with regard to the use of neural resources. Neural efficiency has been discussed in relation to intelligence, where it has been proposed that more intelligent individuals show lower BOLD signal while performing cognitive tasks (Neubauer & Fink, 2009). In this study, there were no group differences in psychometrically evaluated intelligence. Neural efficiency is often observed in tasks of low or moderate difficulty and predominantly in brain regions of the frontal cortex (Neubauer & Fink, 2009). Both of these prerequisites are present in this study. The efficiency of AP processing might be related to the automatic retrieval of the pitch-label association which presumably occurs immediately after the pitch is encoded (Itoh et al., 2005). This process is often described as effortless (Deutsch, 2013). RP requires more processing steps because after the encoding, the pitch needs to be compared to a previous pitch held in working memory and subsequently, the exact interval between those two pitches needs to be determined. One might speculate that the presumed neural efficiency of AP processing could be a reason for its continued existence throughout human evolution despite its negligible role in music and speech perception (McDermott & Oxenham, 2008). On the other hand, it could also be argued that AP musicians did not use the IFG and preSMA at all during Labeling, and thus, the notion of more efficient neural processing might be misplaced, as AP musicians might have used different brain regions than RP musicians and not the same regions more efficiently (see Neubauer & Fink, 2009). Following this line of reasoning, AP musicians may have relied on different cognitive processes during Labeling than RP musicians. However, there are two lines of evidence that speak against the AP-specific use of fundamentally different neural resources in the Labeling condition: First, from the unthresholded statistical map displaying differences in BOLD signal between Listening and

Labeling within AP musicians, one can observe that AP musicians did, to some extent, activate the bilateral IFG and the preSMA more during Labeling than during Listening (see https://neurovault.org/images/117517/). Thus, they actually used the same, or at least similar, brain regions as RP musicians during Labeling. Second, recent behavioral studies have demonstrated that AP processing and higher cognitive functions (e.g., working memory) are more closely related than previously thought (Van Hedger, Heald, Koch, et al., 2015; Van Hedger & Nusbaum, 2018). Hence, it is possible that AP processing is not completely independent of higher cognitive functions but relies less on them than RP processing.

During Listening, the AP musicians showed larger BOLD signals than RP musicians in the right planum temporale. We observed this increase exclusively with the ROI analysis, so the effect seems to be spatially restricted and too subtle to be detected by analyses employing a conservative correction for multiple comparisons. As described in the introduction, the planum temporale has been associated with AP processing from the very beginning of neuroscientific AP research (Schlaug et al., 1995). It is part of the non-primary auditory cortex and has an important role in the processing of a diverse range of sounds (Griffiths & Warren, 2002). In this study, the increase in signal was restricted to the right hemisphere. This finding is consistent with previous studies reporting anatomical differences in AP musicians in the right planum temporale (Keenan et al., 2001; Wengenroth et al., 2014; Wilson et al., 2009) and with an influential theory on the importance of the right hemispheric auditory cortex in music processing (Zatorre, Belin, & Penhune, 2002). However, its exact role in AP processing is still unclear. With regard to auditory processing in general, it has been proposed that the planum temporale matches incoming auditory information with information that is stored in templates which are not located in the planum temporale itself (Griffiths & Warren, 2002). According to the two-component model, AP musicians possess long-term representations of pitches associated with meaningful labels. These representations could well be characterized as internal templates to which incoming information is matched (Levitin, 1994; Levitin & Rogers, 2005). Therefore, we propose that in AP musicians, incoming auditory information, more precisely the extracted pitch information, is matched with these internal pitch templates by computations performed in the right planum temporale. The templates themselves could be represented in more anterior regions of the right temporal lobe which are implicated in semantic memory (Binder & Desai, 2011). In line with this idea, two recent studies investigating AP musicians have found evidence for differential structural and functional connectivity along the right-hemispheric ventral stream of auditory processing, especially in the planum polare which is located immediately anterior to Heschl's gyrus (Kim & Knösche, 2016, 2017). Thus, it will be

interesting for future studies trying to consolidate the findings of AP-specific alterations in posterior and anterior secondary auditory cortices.

In contrast to the previously described PET study, we did not find group differences in the posterior DLPFC during Listening. In the PET study, the involvement of the DLPFC was attributed to the automatic retrieval of the pitch-label association in AP musicians (Zatorre et al., 1998). The current results do not support this interpretation. In both groups, we observed bilateral DLPFC BOLD signal increases during Labeling. These increases were accompanied by higher BOLD signal in the bilateral IPS, again in both groups. Both the DLPFC and the IPS are important parts of a network strongly linked to top-down attentional control (Corbetta & Shulman, 2002). Therefore, it is possible that the DLPFC involvement is related to unspecific attentional processes rather than the specific retrieval of the pitch-label association.

In conclusion, the current results indicate a possible involvement of working memory, language-related processes, and auditory imagery in RP processing, mediated by the bilateral IFG and the preSMA. AP musicians do not show BOLD signal increases in the IFG and the preSMA during Labeling. At the same time, AP musicians label the tones with a higher accuracy. This suggests that AP might be an example of neural efficiency, which is characterized by higher behavioral performance in combination with a lower use of neural resources. Using MVPA, we detected differential BOLD signal patterns in the IFG and the preSMA. Therefore, these regions might contain information differentiating AP from RP on a small spatial scale. Finally, during Listening, the AP musicians show a specific signal increase in the right planum temporale, possibly reflecting the matching of pitch information with internal templates. Taken together, AP and RP musicians show diverging frontal activations during Labeling and, more subtly, differences in right auditory activation during Listening. The results of this study do not support the previously reported importance of the posterior DLPFC in associating a pitch with its label.

Chapter 5 General Discussion

The main aim of this thesis was to elucidate the neural bases of pitch processing in AP and RP musicians. More specifically, we focused on the ability to label pitches, which differentiates between musicians with and without AP. Based on the data collected in two experiments using EEG and task-based fMRI, we gained new insights into what distinguishes the neural processing of pitches in AP musicians from the processing in RP musicians.

In the first article of this thesis, we partially replicated the findings of a seminal study on the electrophysiological correlates of AP and RP (Itoh et al., 2005). We did not find evidence for an AP-specific ERP component called *AP negativity* during simple tone listening. However, we found ERP components related to RP similar to the original study. In contrast to the original study, the elicitation of RP-related ERP components was not gradually dependent on AP ability, but we found the largest differences between participants with middle to high levels of AP and participants with low levels of AP (i.e. RP musicians). These results are consistent with previous studies reporting an absence of evidence for group differences in early ERP components (e.g., Greber et al., 2018).

In article 2 of this thesis, using a "brain decoding" framework, we found that AP musicians showed more similar neural representations of listening and labeling than RP musicians during late stages of pitch processing. The higher representational similarity was present in theta and beta oscillations. The complementary univariate analyses we performed were less sensitive in detecting oscillatory group differences. We concluded that AP and RP musicians primarily differ in stages of pitch processing associated with cognition, rather than earlier stages associated with perception.

Article 3 described an fMRI experiment aiming at localizing differences in brain activation between AP and RP musicians during pitch processing. We found lower activation in AP compared to RP musicians during pitch labeling in the inferior frontal gyrus and the presupplementary motor area. Since AP musicians also showed better performance at the behavioral level, we suggest that AP might be an example for neural efficiency. Using a specific form of MVPA, searchlight analysis, we found that activation patterns in the inferior frontal gyrus and the presupplementary motor area contained information distinguishing AP from RP musicians. Finally, ROI analysis revealed that while listening to tones, AP musicians showed higher activation in the right planum temporale.

5.1 Distinct Population, Continuum, or Both?

Before turning to the integrated discussion of the EEG and fMRI findings, first, some characteristics of our large sample of AP and RP musicians are discussed in more detail. This thesis was conducted within a larger project, in which we recruited more than 100 musicians with and without AP. The participants were categorized into AP and RP musicians primarily based on self-report. In addition, a tone-naming test confirmed the presence of AP in self-reported AP musicians. We did not re-categorize RP musicians that insisted on not possessing AP, even if they performed exceptionally well in the tone-naming test (see also 4.3.1). Categorizing based on both self-report and tone-naming test is advantageous because there is no need for an arbitrary cut-off score. Using this procedure, we found a bimodal distribution of tone-naming scores among our participants. AP musicians clustered around a high performance level (~75% correct identification of pitches) and RP musicians clustered slightly above the chance level of 8.3% (see the left panel of Figure 1 in Chapter 4 for a visualization). The bimodality of the distribution implies two separate populations for AP and RP musicians. Distinct populations have also been suggested by previous studies using different tone-naming tests in large samples (Athos et al., 2007; Wengenroth et al., 2014). For example, Athos et al. (2007) tested more than 2,200 participants in the context of a genetic study. The tone-naming scores in this study were clearly bimodally distributed (and not normally as many other cognitive functions such as intelligence), and the authors concluded that AP was a dichotomous ability present only in a distinct group of individuals of the human population (see also Drayna, 2007). Wengenroth et al. (2014) tested 162 musicians using a tone-naming test specifically developed to exclude RP-based identification strategies and also observed a bimodal distribution.

Apart from these larger-scale studies, several smaller studies have found evidence for (or implicitly assumed) a more continuous distribution of tone-naming scores (Baharloo et al., 1998; Bermudez & Zatorre, 2009; Itoh et al., 2005; Loui et al., 2010, 2012; Wilson et al., 2009, 2012). In these studies, it was assumed that there are many intermediate levels of tone-naming proficiency that lie between chance level and near-perfect performance. Indeed, in our sample and in the samples of Athos et al. (2007) and Wengenroth et al. (2014), some participants showed immediate levels of tone naming. Furthermore, despite the bimodality in our sample, the scores of self-reported AP and self-reported RP musicians partially overlapped, which seems to be a common phenomenon when participants are assigned by self-report as opposed to arbitrary cutoffs (e.g., Dooley & Deutsch, 2010).

The question that now arises is how to integrate these seemingly contradictory findings. One aspect worth considering is that pitch labeling assessed in tone-naming tests is highly variable in AP musicians. Performance depends on the current, recent, and early auditory environment (e.g., listening experiences, musical training, and language exposure), stimulus-related features, and task instructions (see 1.3.2). Qualitative research on variability in tone-naming has identified similar factors influencing performance, e.g., time away from the primary instrument or exposure to a different tuning standard (West Marvin, VanderStel, & Siu, 2019). Moreover, RP musicians also show considerable variability in tone naming. For example, *Quasi-AP* musicians as characterized by Bachem (1937) use an internal reference (e.g., 440 Hz) in combination with their RP ability to label pitches in tone-naming tests. It is conceivable that these individuals are very proficient RP musicians rather than AP musicians. Given the high variability in tone naming in both AP and RP musicians, it might well be that there exists a continuum of tone-naming proficiency *within* each of the two distinct populations of AP and RP musicians. For example, if tone naming is normally distributed within each of the populations, a primarily bimodal distribution over both populations with some intermediate performers would be expected. Additionally, unspecific factors such as inattentiveness or fatigue might add noise to the underlying distributions, making both distributions wider and more likely to overlap. However, one must also consider the sampling strategy employed in a particular study, e.g., voluntary sampling, convenience sampling, or random sampling. For example, in the project this thesis was conducted in, we used a voluntary sampling strategy which has likely biased the distribution of tone-naming scores towards bimodality (Bermudez & Zatorre, 2009). Thus, future large-scale studies on AP should try to employ a random sampling strategy to advance the discussion on the distribution of AP ability.

5.2 AP and Working Memory

Moving on from the characteristics of our sample, the next paragraphs discuss this thesis' implications for the perceptual and cognitive functions involved in AP and RP. Auditory working memory represents the cognitive function which was most consistently discussed across the three articles of this thesis. In the first article describing the replication study of Itoh et al. (2005), we confirmed that two ERP components related to working memory were elicited less strongly in individuals with high levels of AP than in RP musicians, consistent with a number of early ERP studies on AP (e.g., Klein et al., 1984). In the second article, we found that AP musicians showed lower theta power during late stages of pitch processing in the labeling condition. Both theta oscillations and later processing stages are associated with higher cognitive functions and especially working memory (Albouy et al., 2017; Koelsch & Siebel,

2005). In the third article describing the fMRI experiment, we found lower activation in AP musicians during labeling in the inferior frontal gyrus, pars opercularis, a brain region repeatedly associated with working memory (Schulze et al., 2018). One interpretation for these findings is that AP musicians do not rely on auditory working memory during pitch labeling. Because the (verbal) label is part of the long-term representation of the pitch, it can be directly retrieved from semantic memory (Hsieh & Saberi, 2008a). A different interpretation is that AP musicians essentially use the same working memory processes as RP musicians, but as experts in labeling pitches, AP musicians use these processes more efficiently (see also the paragraph on neural efficiency in 4.5).

Although the first interpretation has usually been preferred by previous studies (e.g., Klein et al., 1984; Zatorre et al., 1998), evidence in favor of the second interpretation comes from more recent behavioral studies suggesting that AP and working memory are more intertwined than previously thought (Deutsch & Dooley, 2013; Hutka & Alain, 2015; Van Hedger, Heald, Koch, et al., 2015). First evidence for a stronger involvement of working memory in AP perception came from a study demonstrating that musicians with AP have a larger auditory working memory capacity than musicians without AP (Deutsch & Dooley, 2013). This is surprising because if AP processing was independent of working memory, one would not expect such a relationship. Furthermore, in individuals without AP, working memory capacity explains interindividual differences in learning the associations between pitches and their labels (Van Hedger, Heald, Koch, et al., 2015).

Considering the close association between working memory and RP, the relationship between AP and RP could be stronger than widely assumed and might be mediated by auditory working memory mechanisms. Interestingly, the hypothesis of a close relationship between AP and RP has been stated before, e.g., to explain variability in the elicitation of the P300 component within a group of AP musicians (Bischoff Renninger et al., 2003). Furthermore, tone-naming proficiency is positively associated with the performance in an interval identification test which essentially probes RP ability (Dooley & Deutsch, 2011), a finding we were able to replicate in article 2 of this thesis (see 3.4.2.). Taken together, it could be interesting to further investigate to what extent (and in which tasks) AP musicians also rely on RP processing, and how these two abilities interact with auditory working memory processes.

5.3 AP and Planum Temporale

After having outlined the relevance of this thesis' findings for perceptual and cognitive functions involved in AP and RP processing, in the following, the findings are discussed with

regard to the neural bases of AP. The very first brain region that has been found to be associated with AP was the planum temporale (Schlaug et al., 1995). As part of secondary auditory cortex located immediately posterior to Heschl's gyrus on the superior temporal plane, the planum temporale takes part in the processing of a variety of sounds from noise to musical sounds to environmental and speech sounds. It has been proposed that the planum temporale acts as "computational hub" which segregates auditory objects from the constantly incoming acoustic stimulation and matches these objects to acquired long-term representations. This in turn enables the identification of auditory objects (Griffiths & Warren, 2002). Originally, the (left) planum temporale was thought of as being solely a processor of speech sounds, consistent with its location in the center of Wernicke's area (Wise et al., 2001) and its prominent leftward asymmetry (left larger than right) present in the majority of humans (Geschwind & Levitsky, 1968).

In their seminal study, Schlaug et al. (1995) found an exaggerated leftward asymmetry in AP musicians compared to RP musicians and non-musicians. A subsequent study on planum temporale asymmetry in AP found that the stronger leftward asymmetry was likely driven by a smaller right-sided (rather than a larger left-sided) planum temporale in AP musicians (Keenan et al., 2001). This result has been replicated in more recent anatomical studies on AP (Wengenroth et al., 2014; Wilson et al., 2009). An association of the right planum temporale with AP fits well with our findings from the ROI analysis of the fMRI data in article 3, where we found increased activation specifically in the right-sided, but not the left-sided, planum temporale in AP compared to RP musicians (see also Wengenroth et al., 2014). The fact that this group difference in activation was observed during listening suggests that it might reflect an automatic labeling process that occurs when AP musicians are confronted with sounds that evoke a specific pitch (Itoh et al., 2005). As outlined in the discussion section of article 3, the right planum temporale might match the extracted pitch information from incoming sounds to the pitch templates in semantic memory, which themselves might be represented in more anterior parts of the temporal lobe (see 4.5). It would now be of interest how the Heschl's gyrus, which presumably extracts the pitch information, works in concert with the planum temporale and semantic memory systems to enable the automatic labeling of incoming sounds. Especially because both Heschl's gyrus (McKetton et al., 2019; Wengenroth et al., 2014) and anterior parts of the temporal lobe (Kim & Knösche, 2016, 2017) have been previously implicated in AP processing. Here, an investigation of task-based functional connectivity during pitch processing and, to some extent, analyses of intrinsic functional and structural connectivity could provide valuable insights into how these brain regions interact in AP processing.

5.4 AP and Autism Spectrum Disorders

Anecdotally, AP ability has been frequently associated with autism spectrum disorders, henceforth called autism (e.g., Sacks, 1995), even though the empirical evidence for an association is relatively sparse. Autism is a neurodevelopmental disorder typically characterized by deficits in social communication and interaction, rigid behaviors, interests, and activities (American Psychiatric Association, 2013). Individuals with autism have a higher prevalence of AP than individuals in the general population (Mottron et al., 2013), and there is also evidence for enhanced pitch memory and pitch labeling in autism (Heaton, 2003). Conversely, it has been shown that AP possessors have more autistic traits than non-possessors (Brown et al., 2003; Dohn, Garza-Villarreal, Heaton, & Vuust, 2012). On the neural level, AP and autism are characterized by altered neural connectivity, both functionally and structurally (Brauchli et al., 2019; Courchesne & Pierce, 2005; Jäncke et al., 2012; Loui et al., 2010, 2012; Supekar et al., 2013). A recent EEG study concluded that there are both commonalities and differences in the neural characterization of AP and autism (Wenhart, Bethlehem, et al., 2019).

In article 2 of this thesis, we investigated the representational similarity of experimental conditions in AP and RP musicians. A similar approach has been previously employed in a sample of individuals with autism and typically developing controls (Uddin et al., 2015). The authors found that individuals with autism showed a "neurophysiological inflexibility", characterized by higher representational similarity of different conditions (task vs. rest) than typically developing controls (Uddin et al., 2015). This finding is analogous to the higher representational similarity of listening and labeling we found in AP musicians. Thus, the results of article 2 might hint at a further neural commonality as both AP musicians and individuals with autism show rigid neural processing during different experimental conditions. Although this thesis' contribution to the topic is admittedly modest, the relationship between AP and autism and its neural underpinnings continues to be an active area of research (Wenhart & Altenmüller, 2019; Wenhart, Hwang, & Altenmüller, 2019).

5.5 Future Directions

As noted in the introduction, the cognitive neuroscience of AP is still in its early stages as only a small number of studies have been conducted, and most of these studies have had small sample sizes. Put differently, there is still much more to learn about the neural bases of this puzzling ability. The articles in this thesis contribute towards a better understanding by assessing the robustness of highly influential previous findings and by providing novel insights into the electrophysiological aspects and the functional anatomy of AP (and RP) processing.

However, more research is needed to answer the many open questions on the neural mechanisms underlying the phenomenon. In the following, directions for future research are provided that might enable the cognitive neuroscience of AP to transition from its current early to a more mature stage.

5.5.1 AP and Models of Auditory Processing

A first potential direction that could advance the understanding of AP is to embed findings within the context of neural models of auditory processing. As the explicit labeling of a pitch is essentially the final step in a long cascade of processing steps involving the whole auditory system and a network of higher-order brain areas in the temporal, parietal, and frontal lobes, it might be useful to think about AP processing as a specific variant of auditory processing. We have recently proposed that prominent dual-stream models of auditory processing (Friederici, 2011; Hickok & Poeppel, 2007; Rauschecker & Scott, 2009; Rauschecker & Tian, 2000) might provide a suitable framework for guiding future research on the cortical processing of pitches in AP musicians (Leipold, Greber, & Elmer, 2019). Apart from the dual-stream models which are primarily concerned with cortical processing of sounds, a different framework might be provided by models considering the subcortical processing of sounds along the auditory pathway from the cochlea along the brainstem nuclei, midbrain, and thalamus (e.g., McLachlan & Wilson, 2010). For example, there has been no published study to date on subcortical auditory processing in AP compared to RP musicians. In contrast, previous studies investigating a potential peripheral basis of AP have not found evidence for differential peripheral processing of pitches in AP (McKetton, Purcell, Stone, Grahn, & Bergevin, 2018).

5.5.2 Reproducibility and Collaborations

As shown by the replication study described in Chapter 2 and by a further EEG study on AP (Greber et al., 2018), the robustness of previous findings on AP may not be as high as desired. To ensure a high robustness of our own findings, we invested considerable time and resources into the recruitment of a sufficient number of participants. However, it would be more efficient if the requirements for recruiting and measuring a large sample were shared between multiple research groups studying AP and RP processing. Research using larger samples has a higher probability of finding true and replicable effects and thus, collaborations would foster reproducibility. As a first step in this direction, we uploaded the raw EEG data used in articles 1 and 2 (and the materials from all three articles) to the Open Science Framework (https://dx.doi.org/10.17605/OSF.IO/7QXJS).

5.6 Conclusion

To conclude, this thesis advances our understanding of neural differences in pitch processing between AP musicians and RP musicians. The neurophysiological timing and neuroanatomical localization of group differences during pitch labeling suggest a differential involvement of cognitive processes, such as auditory working memory, in AP and RP processing. The localization of more subtle differences during listening also points to perceptual differences between AP and RP musicians, possibly linked to the automatic matching of extracted pitch information to internal pitch templates in AP musicians. More generally, the findings of this thesis highlight the need for studies on large samples, which can be performed more efficiently through collaborative efforts across multiple research groups. Future work on AP and RP processing will benefit from the integration of findings within general-purpose models on the neurobiology of auditory processing.

Chapter 6 References

Akiva-Kabiri, L., & Henik, A. (2012). A Unique Asymmetrical Stroop Effect in Absolute Pitch Possessors. *Experimental Psychology*, *59*(5), 272–278. https://doi.org/10.1027/1618-3169/a000153

Albers, C., & Lakens, D. (2018). When power analyses based on pilot data are biased: inaccurate effect size estimators and follow-up bias. *Journal of Experimental Social Psychology*, *74*, 187–195. https://doi.org/10.1016/J.JESP.2017.09.004

Albouy, P., Mattout, J., Bouet, R., Maby, E., Sanchez, G., Aguera, P.-E., … Tillmann, B. (2013). Impaired pitch perception and memory in congenital amusia: the deficit starts in the auditory cortex. *Brain*, *136*(5), 1639–1661. https://doi.org/10.1093/brain/awt082

Albouy, P., Weiss, A., Baillet, S., & Zatorre, R. J. (2017). Selective entrainment of theta oscillations in the dorsal stream causally enhances auditory working memory performance. *Neuron*, *94*(1), 193-206.e5. https://doi.org/10.1016/J.NEURON.2017.03.015

American Psychiatric Association. (2013). Neurodevelopmental Disorders. In *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*. American Psychiatric Publishing. https://doi.org/10.1176/appi.books.9780890425596.dsm01

Anderson, S. F., & Maxwell, S. E. (2016). There's more than one way to conduct a replication study: beyond statistical significance. *Psychological Methods*, *21*(1), 1–12. https://doi.org/10.1037/met0000051

Annett, M. (1970). A classification of hand preference by association analysis. *British Journal of Psychology*, *61*(3), 303–321. https://doi.org/10.1111/j.2044-8295.1970.tb01248.x

Athos, E. A., Levinson, B., Kistler, A., Zemansky, J., Bostrom, A., Freimer, N. B., & Gitschier, J. (2007). Dichotomy and perceptual distortions in absolute pitch ability. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(37), 14795–14800. https://doi.org/10.1073/pnas.0703868104

Bachem, A. (1937). Various types of absolute pitch. *The Journal of the Acoustical Society of America*, *9*(2), 146–151. https://doi.org/10.1121/1.1915919

Bachem, A. (1940). The genesis of absolute pitch. *The Journal of the Acoustical Society of America*, *11*(4), 434–439. https://doi.org/10.1121/1.1916056

Bachem, A. (1955). Absolute pitch. *The Journal of the Acoustical Society of America*, *27*(6), 1180–1185. https://doi.org/10.1121/1.1908155

Baharloo, S., Johnston, P. A., Service, S. K., Gitschier, J., & Freimer, N. B. (1998). Absolute Pitch: An Approach for Identification of Genetic and Nongenetic Components. *The American Journal of Human Genetics*, *62*(2), 224–231. https://doi.org/10.1086/301704

Baharloo, S., Service, S. K., Risch, N., Gitschier, J., & Freimer, N. B. (2000). Familial aggregation of absolute pitch. *The American Journal of Human Genetics*, *67*(3), 755–758. https://doi.org/10.1086/303057

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*(7604), 452–454. https://doi.org/10.1038/533452a

Begley, C. G., & Ellis, L. M. (2012). Raise standards for preclinical cancer research. *Nature*, *483*(7391), 531–533. https://doi.org/10.1038/483531a

Behroozmand, R., Ibrahim, N., Korzyukov, O., Robin, D. A., & Larson, C. R. (2014). Left-hemisphere activation is associated with enhanced vocal pitch error detection in musicians with absolute pitch. *Brain and Cognition*, *84*(1), 97–108. https://doi.org/10.1016/J.BANDC.2013.11.007

Behroozmand, R., Ibrahim, N., Korzyukov, O., Robin, D. A., & Larson, C. R. (2015). Functional role of delta and theta band oscillations for auditory feedback processing during vocal pitch motor control. *Frontiers in Neuroscience*, *9*, 109. https://doi.org/10.3389/fnins.2015.00109

Bendor, D. (2012). Does a pitch center exist in auditory cortex? *Journal of Neurophysiology*, *107*(3), 743–746. https://doi.org/10.1152/jn.00804.2011

Bermudez, P., Lerch, J. P., Evans, A. C., & Zatorre, R. J. (2009). Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. *Cerebral Cortex*, *19*(7), 1583–1596. https://doi.org/10.1093/cercor/bhn196

Bermudez, P., & Zatorre, R. J. (2005). Conditional associative memory for musical stimuli in nonmusicians: Implications for absolute pitch. *The Journal of Neuroscience*, *25*(34), 7718–7723. https://doi.org/10.1523/jneurosci.1560-05.2005

Bermudez, P., & Zatorre, R. J. (2009). A distribution of absolute pitch ability as revealed by computerized testing. *Music Perception: An Interdisciplinary Journal*, *27*(2), 89–101. https://doi.org/10.1525/mp.2009.27.2.89

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, *15*(11), 527–536. https://doi.org/10.1016/J.TICS.2011.10.001

Bischoff Renninger, L., Granot, R. I., & Donchin, E. (2003). Absolute pitch and the P300 component of the event-related potential: An exploration of variables that may account for individual differences. *Music Perception: An Interdisciplinary Journal*, *20*(4), 357–382. https://doi.org/10.1525/mp.2003.20.4.357

Blankertz, B., Lemm, S., Treder, M., Haufe, S., & Müller, K. R. (2011). Single-trial analysis and classification of ERP components - A tutorial. *NeuroImage*, *56*(2), 814–825. https://doi.org/10.1016/j.neuroimage.2010.06.048

Boekel, W., Wagenmakers, E.-J., Belay, L., Verhagen, J., Brown, S., & Forstmann, B. U. (2015). A purely confirmatory replication study of structural brain-behavior correlations. *Cortex*, *66*, 115–133. https://doi.org/10.1016/J.CORTEX.2014.11.019

Brauchli, C., Leipold, S., & Jäncke, L. (2019). Univariate and multivariate analyses of functional networks in absolute pitch. *NeuroImage*, *189*, 241–247. https://doi.org/10.1016/J.NEUROIMAGE.2019.01.021

Brown, W. A., Cammuso, K., Sachs, H., Winklosky, B., Mullane, J., Bernier, R., … Folstein, S. E. (2003). Autism-Related Language, Personality, and Cognition in People with Absolute Pitch: Results of a Preliminary Study. *Journal of Autism and Developmental Disorders*, *33*(2), 163–167. https://doi.org/10.1023/A:1022987309913

Buchanan, T. W., Lutz, K., Mirzazade, S., Specht, K., Shah, N. J., Zilles, K., & Jäncke, L. (2000). Recognition of emotional prosody and verbal components of spoken language: An fMRI study. *Cognitive Brain Research*, *9*(3), 227–238. https://doi.org/10.1016/S0926-6410(99)00060-9

Burkhard, A., Elmer, S., & Jäncke, L. (2019). Early tone categorization in absolute pitch musicians is subserved by the right-sided perisylvian brain. *Scientific Reports*, *9*(1), 1419. https://doi.org/10.1038/s41598-018-38273-0

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. https://doi.org/10.1038/nrn3475

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., … Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, *351*(6280), 1433–1436. https://doi.org/10.1126/science.aaf0918

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., … Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, 1. https://doi.org/10.1038/s41562-018-0399-z

Catani, M., Mesulam, M. M., Jakobsen, E., Malik, F., Martersteck, A., Wieneke, C., … Rogalski, E. (2013). A novel frontal pathway underlies verbal fluency in primary progressive aphasia. *Brain*, *136*(8), 2619–2628. https://doi.org/10.1093/brain/awt163

Chavarria-Soley, G. (2016). Absolute pitch in Costa Rica: Distribution of pitch identification ability and implications for its genetic basis. *The Journal of the Acoustical Society of America*, *140*(2), 891–897. https://doi.org/10.1121/1.4960569

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, *17*(3), 455–462. https://doi.org/10.1038/nn.3635

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.

https://doi.org/10.1037/0033-2909.112.1.155
Cohen, M. X. (2014). *Analyzing neural time series data: theory and practice*. MIT Press. Retrieved from https://mitpress.mit.edu/books/analyzing-neural-time-series-data
Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201–215. https://doi.org/10.1038/nrn755
Courchesne, E., & Pierce, K. (2005). Why the frontal cortex in autism might be talking only to itself: local over-connectivity but long-distance disconnection. *Current Opinion in Neurobiology*, *15*(2), 225–230. https://doi.org/10.1016/J.CONB.2005.03.001
Craddock, M. (2018). craddm/eegUtils: eegUtils. https://doi.org/10.5281/ZENODO.1292902
Cramer, A. O. J., van Ravenzwaaij, D., Matzke, D., Steingroever, H., Wetzels, R., Grasman, R. P. P. P., … Wagenmakers, E.-J. (2016). Hidden multiplicity in exploratory multiway ANOVA: prevalence and remedies. *Psychonomic Bulletin & Review*, *23*(2), 640–647. https://doi.org/10.3758/s13423-015-0913-5
Crouzet, S. M., Busch, N. A., & Ohla, K. (2015). Taste Quality Decoding Parallels Taste Sensations. *Current Biology*, *25*(7), 890–896. https://doi.org/10.1016/J.CUB.2015.01.057
Crummer, G. C., Walton, J. P., Wayman, J. W., Hantz, E. C., & Frisina, R. D. (1994). Neural processing of musical timbre by musicians, nonmusicians, and musicians possessing absolute pitch. *The Journal of the Acoustical Society of America*, *95*(5), 2720–2727. https://doi.org/10.1121/1.409840
Deutsch, D. (1986). A Musical Paradox. *Music Perception: An Interdisciplinary Journal*, *3*(3), 275–280. https://doi.org/10.2307/40285337
Deutsch, D. (2002). The Puzzle of Absolute Pitch. *Current Directions in Psychological Science*, *11*(6), 200–204. https://doi.org/10.1111/1467-8721.00200
Deutsch, D. (2013). Absolute pitch. In *The Psychology of Music* (pp. 141–182). Elsevier. https://doi.org/10.1016/B978-0-12-381460-9.00005-5
Deutsch, D., & Dooley, K. (2013). Absolute pitch is associated with a large auditory digit span: A clue to its genesis. *The Journal of the Acoustical Society of America*, *133*(4), 1859–1861. https://doi.org/10.1121/1.4792217
Deutsch, D., Dooley, K., Henthorn, T., & Head, B. (2009). Absolute pitch among students in an American music conservatory: Association with tone language fluency. *The Journal of the Acoustical Society of America*, *125*(4), 2398–2403. https://doi.org/10.1121/1.3081389
Deutsch, D., Henthorn, T., & Dolson, M. (1999). Absolute pitch is demonstrated in speakers of tone languages. *The Journal of the Acoustical Society of America*, *106*(4), 2267–2267. https://doi.org/10.1121/1.427738
Deutsch, D., Henthorn, T., & Dolson, M. (2004). Absolute pitch, speech, and tone language: Some experiments and a proposed framework. *Music Perception: An Interdisciplinary Journal*, *21*(3), 339–356. https://doi.org/10.1525/mp.2004.21.3.339
Deutsch, D., Henthorn, T., Marvin, E. W., & Xu, H. (2006). Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period. *The Journal of the Acoustical Society of America*, *119*(2), 719–722. https://doi.org/10.1121/1.2151799
Dienes, Z. (2011). Bayesian versus orthodox statistics: which side are you on? *Perspectives on Psychological Science*, *6*(3), 274–290. https://doi.org/10.1177/1745691611406920
Dohn, A., Garza-Villarreal, E. A., Chakravarty, M. M., Hansen, M., Lerch, J. P., & Vuust, P. (2015). Gray- and white-matter anatomy of absolute pitch possessors. *Cerebral Cortex*, *25*(5), 1379–1388. https://doi.org/10.1093/cercor/bht334
Dohn, A., Garza-Villarreal, E. A., Heaton, P., & Vuust, P. (2012). Do musicians with perfect pitch have more autism traits than musicians without perfect pitch? An empirical study. *PLoS ONE*, *7*(5), e37961. https://doi.org/10.1371/journal.pone.0037961
Dohn, A., Garza-Villarreal, E. A., Ribe, L. R., Wallentin, M., & Vuust, P. (2014). Musical Activity Tunes Up Absolute Pitch Ability. *Music Perception: An Interdisciplinary Journal*, *31*(4), 359–371. https://doi.org/10.1525/mp.2014.31.4.359
Dooley, K., & Deutsch, D. (2010). Absolute pitch correlates with high performance on musical dictation. *The Journal of the Acoustical Society of America*, *128*(2), 890–893. https://doi.org/10.1121/1.3458848

Dooley, K., & Deutsch, D. (2011). Absolute pitch correlates with high performance on interval naming tasks. *The Journal of the Acoustical Society of America*, *130*(6), 4097–4104. https://doi.org/10.1121/1.3652861

Drayna, D. T. (2007). Absolute pitch: a special group of ears. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(37), 14549–14550. https://doi.org/10.1073/pnas.0707287104

Drayna, D. T., Manichaikul, A., de Lange, M., Snieder, H., & Spector, T. (2001). Genetic correlates of musical pitch recognition in humans. *Science*, *291*(5510), 1969–1972. https://doi.org/10.1126/science.291.5510.1969

Eden, G. F., Joseph, J. E., Brown, H. E., Brown, C. P., & Zeffiro, T. A. (1999). Utilizing hemodynamic delay and dispersion to detect fMRI signal change without auditory interference: The behavior interleaved gradients technique. *Magnetic Resonance in Medicine*, *41*(1), 13–20. https://doi.org/10.1002/(SICI)1522-2594(199901)41:1<13::AID-MRM4>3.0.CO;2-T

Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(28), 7900–7905. https://doi.org/10.1073/pnas.1602413113

Elmer, S., Sollberger, S., Meyer, M., & Jäncke, L. (2013). An empirical reevaluation of absolute pitch: behavioral and electrophysiological measurements. *Journal of Cognitive Neuroscience*, *25*(10), 1736–1753. https://doi.org/10.1162/jocn_a_00410

Engel, A. K., & Fries, P. (2010). Beta-band oscillations — signalling the status quo? *Current Opinion in Neurobiology*, *20*(2), 156–165. https://doi.org/10.1016/J.CONB.2010.02.015

Etzel, J. A., Zacks, J. M., & Braver, T. S. (2013). Searchlight analysis: Promise, pitfalls, and potential. *NeuroImage*, *78*, 261–269. https://doi.org/10.1016/j.neuroimage.2013.03.041

Foster, N. E. V., & Zatorre, R. J. (2010a). A role for the intraparietal sulcus in transforming musical pitch information. *Cerebral Cortex*, *20*(6), 1350–1359. https://doi.org/10.1093/cercor/bhp199

Foster, N. E. V., & Zatorre, R. J. (2010b). Cortical structure predicts success in performing musical transformation judgments. *NeuroImage*, *53*(1), 26–36. https://doi.org/10.1016/J.NEUROIMAGE.2010.06.042

Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *Physiological Reviews*, *91*(4), 1357–1392. https://doi.org/10.1152/physrev.00006.2011

Frieler, K., Fischinger, T., Schlemmer, K., Lothwesen, K., Jakubowski, K., & Müllensiefen, D. (2013). Absolute memory for pitch: A comparative replication of Levitin's 1994 study in six European labs. *Musicae Scientiae*, *17*(3), 334–349. https://doi.org/10.1177/1029864913493802

Friston, K. J., Stephan, K. M., Heather, J. D., Frith, C. D., Ioannides, A. A., Liu, L. C., … Frackowiak, R. S. J. (1996). A Multivariate Analysis of Evoked Responses in EEG and MEG Data. *NeuroImage*, *3*(3), 167–174. https://doi.org/10.1006/NIMG.1996.0018

Fujioka, T., Ross, B., Kakigi, R., Pantev, C., & Trainor, L. J. (2006). One year of musical training affects development of auditory cortical-evoked fields in young children. *Brain*, *129*(10), 2593–2608. https://doi.org/10.1093/brain/awl247

Gaab, N., Gaser, C., Zaehle, T., Jäncke, L., & Schlaug, G. (2003). Functional anatomy of pitch memory—an fMRI study with sparse temporal sampling. *NeuroImage*, *19*(4), 1417–1426. https://doi.org/10.1016/S1053-8119(03)00224-6

García-Larrea, L., & Cézanne-Bert, G. (1998). P3, positive slow wave and working memory load: a study on the functional correlates of slow wave activity. *Electroencephalography and Clinical Neurophysiology*, *108*(3), 260–273. https://doi.org/10.1016/S0168-5597(97)00085-3

Gelman, A., & Stern, H. (2006). The difference between "significant" and "not significant" is not itself statistically significant. *The American Statistician*, *60*(4), 328–331. https://doi.org/10.1198/000313006X152649

Geschwind, N., & Levitsky, W. (1968). Human brain: left-right asymmetries in temporal speech region. *Science*, *161*(3837), 186–187.

https://doi.org/10.1126/science.161.3837.186

Gordon, E. E. (1989). *Advanced Measures of Music Audiation*. GIA Publications.

Gorgolewski, K. J., Varoquaux, G., Rivera, G., Schwarz, Y., Ghosh, S. S., Maumet, C., … Margulies, D. S. (2015). NeuroVault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *Frontiers in Neuroinformatics*, *9*, 8. https://doi.org/10.3389/fninf.2015.00008

Gramfort, A., Papadopoulo, T., Olivi, E., & Clerc, M. (2010). OpenMEEG: opensource software for quasistatic bioelectromagnetics. *BioMedical Engineering OnLine*, *9*(1), 45. https://doi.org/10.1186/1475-925X-9-45

Greber, M., Rogenmoser, L., Elmer, S., & Jäncke, L. (2018). Electrophysiological correlates of absolute pitch in a passive auditory oddball paradigm: A direct replication attempt. *ENeuro*, *5*(6), ENEURO.0333-18.2018. https://doi.org/10.1523/ENEURO.0333-18.2018

Gregersen, P. K., Kowalsky, E., Kohn, N., & Marvin, E. W. (1999). Absolute pitch: Prevalence, ethnic variation, and estimation of the genetic component. *The American Journal of Human Genetics*, *65*(3), 911–913. https://doi.org/10.1086/302541

Gregersen, P. K., Kowalsky, E., Kohn, N., & Marvin, E. W. (2001). Early childhood music education and predisposition to absolute pitch: Teasing apart genes and environment. *American Journal of Medical Genetics*, *98*(3), 280–282. https://doi.org/10.1002/1096-8628(20010122)98:3<280::AID-AJMG1083>3.0.CO;2-6

Gregersen, P. K., Kowalsky, E., Lee, A., Baron-Cohen, S., Fisher, S. E., Asher, J. E., … Li, W. (2013). Absolute pitch exhibits phenotypic and genetic overlap with synesthesia. *Human Molecular Genetics*, *22*(10), 2097–2104. https://doi.org/10.1093/hmg/ddt059

Griffiths, T. D., & Hall, D. A. (2012). Mapping pitch representation in neural ensembles with fMRI. *The Journal of Neuroscience*, *32*(39), 13343–13347. https://doi.org/10.1523/JNEUROSCI.3813-12.2012

Griffiths, T. D., & Warren, J. D. (2002). The planum temporale as a computational hub. *Trends in Neurosciences*, *25*(7), 348–353. https://doi.org/10.1016/S0166-2236(02)02191-4

Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2016). Decoding dynamic brain patterns from evoked responses: A tutorial on multivariate pattern analysis applied to time series neuroimaging data. *Journal of Cognitive Neuroscience*, *29*(4), 677–697. https://doi.org/10.1162/jocn_a_01068

Halpern, A. R. (1989). Memory for the absolute pitch of familiar songs. *Memory & Cognition*, *17*(5), 572–581. https://doi.org/10.3758/BF03197080

Halsey, L. G., Curran-Everett, D., Vowler, S. L., & Drummond, G. B. (2015). The fickle P value generates irreproducible results. *Nature Methods*, *12*(3), 179–185. https://doi.org/10.1038/nmeth.3288

Hantz, E. C., Crummer, G. C., Wayman, J. W., Walton, J. P., & Frisina, R. D. (1992). Effects of musical training and absolute pitch on the neural processing of melodic intervals: a P3 event-related potential study. *Music Perception: An Interdisciplinary Journal*, *10*(1), 25–42. https://doi.org/10.2307/40285536

Hausfeld, L., De Martino, F., Bonte, M., & Formisano, E. (2012). Pattern analysis of EEG responses to speech and voice: Influence of feature grouping. *NeuroImage*, *59*(4), 3641–3651. https://doi.org/10.1016/J.NEUROIMAGE.2011.11.056

Haxby, J. V., Connolly, A. C., & Guntupalli, J. S. (2014). Decoding Neural Representational Spaces Using Multivariate Pattern Analysis. *Annual Review of Neuroscience*, *37*(1), 435–456. https://doi.org/10.1146/annurev-neuro-062012-170325

Heaton, P. (2003). Pitch memory, labelling and disembedding in autism. *Journal of Child Psychology and Psychiatry*, *44*(4), 543–551. https://doi.org/10.1111/1469-7610.00143

Hedger, S. C., Heald, S. L. M., & Nusbaum, H. C. (2013). Absolute pitch may not be so absolute. *Psychological Science*, *24*(8), 1496–1502. https://doi.org/10.1177/0956797612473310

Henson, R. (2007). Efficient experimental design for fMRI. In *Statistical Parametric Mapping: The Analysis of Functional Brain Images* (pp. 193–210). Academic Press London. https://doi.org/10.1016/B978-012372560-8/50015-2

Henthorn, T., & Deutsch, D. (2007). Ethnicity versus early environment: Comment on 'Early

Childhood Music Education and Predisposition to Absolute Pitch: Teasing Apart Genes and Environment' by Peter K. Gregersen, Elena Kowalsky, Nina Kohn, and Elizabeth West Marvin [2000]. *American Journal of Medical Genetics Part A*, *143A*(1), 102–103. https://doi.org/10.1002/ajmg.a.31596

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402. https://doi.org/10.1038/nrn2113

Hsieh, I.-H., & Saberi, K. (2008a). Dissociation of procedural and semantic memory in absolute-pitch processing. *Hearing Research*, *240*(1–2), 73–79. https://doi.org/10.1016/j.heares.2008.01.017

Hsieh, I.-H., & Saberi, K. (2008b). Language-Selective Interference with Long-Term Memory for Musical Pitch. *Acta Acustica United with Acustica*, *94*(4), 588–593. https://doi.org/10.3813/AAA.918068

Hutka, S. A., & Alain, C. (2015). The effects of absolute pitch and tone language on pitch processing and encoding in musicians. *Music Perception: An Interdisciplinary Journal*, *32*(4), 344–354. https://doi.org/10.1525/mp.2015.32.4.344

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Ioannidis, J. P. A. (2008). Why most discovered true associations are inflated. *Epidemiology*. Lippincott Williams & Wilkins. https://doi.org/10.2307/25662607

Itoh, K., Suwazono, S., Arao, H., Miyazaki, K., & Nakada, T. (2005). Electrophysiological correlates of absolute pitch and relative pitch. *Cerebral Cortex*, *15*(6), 760–769. https://doi.org/10.1093/cercor/bhh177

Jakubowski, K., & Müllensiefen, D. (2013). The influence of music-elicited emotions and relative pitch on absolute pitch memory for familiar melodies. *Quarterly Journal of Experimental Psychology*, *66*(7), 1259–1267. https://doi.org/10.1080/17470218.2013.803136

Jäncke, L., Langer, N., & Hänggi, J. (2012). Diminished whole-brain but enhanced peri-sylvian connectivity in absolute pitch musicians. *Journal of Cognitive Neuroscience*, *24*(6), 1447–1461. https://doi.org/10.1162/jocn_a_00227

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press.

Jimura, K., & Poldrack, R. A. (2012). Analyses of regional-average activation and multivoxel pattern information tell complementary stories. *Neuropsychologia*, *50*(4), 544–552. https://doi.org/https://doi.org/10.1016/j.neuropsychologia.2011.11.007

Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., Mckeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, *37*(02), 163–178. https://doi.org/10.1111/1469-8986.3720163

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572

Keenan, J. P., Thangaraj, V., Halpern, A. R., & Schlaug, G. (2001). Absolute pitch and planum temporale. *NeuroImage*, *14*(6), 1402–1408. https://doi.org/10.1006/nimg.2001.0925

Kerr, N. L. (1998). HARKing: hypothesizing after the results are known. *Personality and Social Psychology Review*, *2*(3), 196–217. https://doi.org/10.1207/s15327957pspr0203_4

Kim, S.-G., & Knösche, T. R. (2016). Intracortical myelination in musicians with absolute pitch: Quantitative morphometry using 7-T MRI. *Human Brain Mapping*, *37*(10), 3486–3501. https://doi.org/10.1002/hbm.23254

Kim, S.-G., & Knösche, T. R. (2017). Resting state functional connectivity of the ventral auditory pathway in musicians with absolute pitch. *Human Brain Mapping*, *38*(8), 3899–3916. https://doi.org/10.1002/hbm.23637

King, J. R., Faugeras, F., Gramfort, A., Schurger, A., El Karoui, I., Sitt, J. D., … Dehaene, S. (2013). Single-trial decoding of auditory novelty responses facilitates the detection of residual consciousness. *NeuroImage*, *83*, 726–738. https://doi.org/10.1016/J.NEUROIMAGE.2013.07.013

Klein, M., Coles, M. G. H., & Donchin, E. (1984). People with absolute pitch process tones without producing a P300. *Science*, *223*(4642), 1306–1309. https://doi.org/10.1126/science.223.4642.1306

Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Müller, K., & Gruber, O. (2009). Functional architecture of verbal and tonal working memory: An FMRI study. *Human Brain Mapping*, *30*(3), 859–873. https://doi.org/10.1002/hbm.20550

Koelsch, S., & Siebel, W. A. (2005). Towards a neural basis of music perception. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2005.10.001

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, *11*(8), 599–605. https://doi.org/10.1038/nrn2882

Kriegeskorte, N., & Bandettini, P. A. (2007). Analyzing for information, not activation, to exploit high-resolution fMRI. *NeuroImage*, *38*(4), 649–662. https://doi.org/10.1016/j.neuroimage.2007.02.022

Kriegeskorte, N., Goebel, R., & Bandettini, P. A. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–3868. https://doi.org/10.1073/pnas.0600244103

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4. https://doi.org/10.3389/neuro.06.004.2008

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nature Neuroscience*, *12*(5), 535–540. https://doi.org/10.1038/nn.2303

Lacadie, C. M., Fulbright, R. K., Rajeevan, N., Constable, R. T., & Papademetris, X. (2008). More accurate Talairach coordinates for neuroimaging using non-linear registration. *NeuroImage*, *42*(2), 717–725. https://doi.org/10.1016/J.NEUROIMAGE.2008.04.240

Lee, Y.-S., Turkeltaub, P., Granger, R., & Raizada, R. D. S. (2012). Categorical speech processing in Broca's area: An fMRI study using multivariate pattern-based analysis. *The Journal of Neuroscience*, *32*(11), 3942–3948. https://doi.org/10.1523/JNEUROSCI.3814-11.2012

Lehrl, S. (2005). *Mehrfachwahl-Wortschatz-Intelligenztest MWT-B* (5th ed.). Spitta Verlag.

Lehrl, S., Gallwitz, A., Blaha, L., & Fischer, B. (1991). *Kurztest für allgemeine Basisgrößen der Informationsverarbeitung*. Vless Verlag.

Leipold, S., Brauchli, C., Greber, M., & Jäncke, L. (2019). Absolute and relative pitch processing in the human brain: neural and behavioral evidence. *Brain Structure and Function*, *224*(5), 1723–1738. https://doi.org/10.1007/s00429-019-01872-2

Leipold, S., Greber, M., & Elmer, S. (2019). Perception and Cognition in Absolute Pitch: Distinct yet Inseparable. *The Journal of Neuroscience*, *39*(30), 5839–5841. https://doi.org/10.1523/JNEUROSCI.0653-19.2019

Leipold, S., Oderbolz, C., Greber, M., & Jäncke, L. (2019). A reevaluation of the electrophysiological correlates of absolute pitch and relative pitch: No evidence for an absolute pitch-specific negativity. *International Journal of Psychophysiology*, *137*, 21–31. https://doi.org/10.1016/j.ijpsycho.2018.12.016

Leite, R. B. C., Mota-Rolim, S. A., & Queiroz, C. M. T. (2016). Music Proficiency and Quantification of Absolute Pitch: A Large-Scale Study among Brazilian Musicians. *Frontiers in Neuroscience*, *10*, 447. https://doi.org/10.3389/fnins.2016.00447

Levitin, D. J. (1994). Absolute memory for musical pitch: evidence from the production of learned melodies. *Perception & Psychophysics*, *56*(4), 414–423. https://doi.org/10.3758/bf03206733

Levitin, D. J., & Rogers, S. E. (2005). Absolute pitch: perception, coding, and controversies. *Trends in Cognitive Sciences*, *9*(1), 26–33. https://doi.org/10.1016/j.tics.2004.11.007

Lima, C. F., Krishnan, S., & Scott, S. K. (2016). Roles of supplementary motor areas in auditory processing and auditory imagery. *Trends in Neurosciences*, *39*(8), 527–542. https://doi.org/10.1016/J.TINS.2016.06.003

Loui, P., Li, H. C., Hohmann, A., & Schlaug, G. (2010). Enhanced cortical connectivity in absolute pitch musicians: a model for local hyperconnectivity. *Journal of Cognitive Neuroscience*, *23*(4), 1015–1026. https://doi.org/10.1162/jocn.2010.21500

Loui, P., Zamm, A., & Schlaug, G. (2012). Enhanced functional networks in absolute pitch. *NeuroImage*, *63*(2), 632–640.

https://doi.org/http://dx.doi.org/10.1016/j.neuroimage.2012.07.030
Luck, S. J. (2014). *An introduction to the event-related potential technique* (2nd ed.). MIT Press. Retrieved from https://mitpress.mit.edu/books/introduction-event-related-potential-technique-second-edition
Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology*, *54*(1), 146–157. https://doi.org/10.1111/psyp.12639
Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/https://doi.org/10.1016/j.jneumeth.2007.03.024
McDermott, J. H., & Oxenham, A. J. (2008). Music perception, pitch, and the auditory system. *Current Opinion in Neurobiology*, *18*(4), 452–463. https://doi.org/10.1016/J.CONB.2008.09.005
McKetton, L., DeSimone, K., & Schneider, K. A. (2019). Larger auditory cortical area and broader frequency tuning underlie absolute pitch. *The Journal of Neuroscience*, 1532–18. https://doi.org/10.1523/JNEUROSCI.1532-18.2019
McKetton, L., Purcell, D., Stone, V., Grahn, J., & Bergevin, C. (2018). No otoacoustic evidence for a peripheral basis of absolute pitch. *Hearing Research*, *370*, 201–208. https://doi.org/10.1016/J.HEARES.2018.08.001
McLachlan, N., & Wilson, S. (2010). The central role of recognition in auditory perception: A neurobiological model. *Psychological Review*, *117*(1), 175–196. https://doi.org/10.1037/a0018063
Michel, C. M., & Koenig, T. (2018). EEG microstates as a tool for studying the temporal dynamics of whole-brain neuronal networks: A review. *NeuroImage*, *180*, 577–593. https://doi.org/10.1016/J.NEUROIMAGE.2017.11.062
Michel, C. M., & Murray, M. M. (2012). Towards the utilization of EEG as a brain imaging tool. *NeuroImage*, *61*(2), 371–385. https://doi.org/10.1016/J.NEUROIMAGE.2011.12.039
Miyazaki, K. (1988). Musical pitch identification by absolute pitch possessors. *Perception & Psychophysics*, *44*(6), 501–512. https://doi.org/10.3758/bf03207484
Miyazaki, K. (1989). Absolute pitch identification: Effects of timbre and pitch region. *Music Perception: An Interdisciplinary Journal*, *7*(1), 1–14. https://doi.org/10.2307/40285445
Miyazaki, K. (1990). The speed of musical pitch identification by absolute-pitch possessors. *Music Perception: An Interdisciplinary Journal*, *8*(2), 177–188. https://doi.org/10.2307/40285495
Miyazaki, K., Makomaska, S., & Rakowski, A. (2012). Prevalence of absolute pitch: A comparison between Japanese and Polish music students. *The Journal of the Acoustical Society of America*, *132*(5), 3484–3493. https://doi.org/10.1121/1.4756956
Miyazaki, K., & Rakowski, A. (2002). Recognition of notated melodies by possessors and nonpossessors of absolute pitch. *Perception & Psychophysics*, *64*(8), 1337–1345. https://doi.org/10.3758/BF03194776
Miyazaki, K., Rakowski, A., Makomaska, S., Jiang, C., Tsuzaki, M., Oxenham, A. J., … Lipscomb, S. D. (2018). Absolute Pitch and Relative Pitch in Music Students in the East and the West. *Music Perception: An Interdisciplinary Journal*, *36*(2), 135–155. https://doi.org/10.1525/mp.2018.36.2.135
Mottron, L., Bouvet, L., Bonnel, A., Samson, F., Burack, J. A., Dawson, M., & Heaton, P. (2013). Veridical mapping in the development of exceptional autistic abilities. *Neuroscience & Biobehavioral Reviews*, *37*(2), 209–228. https://doi.org/10.1016/J.NEUBIOREV.2012.11.016
Mumford, J. A. (2012). A power calculation guide for fMRI studies. *Social Cognitive and Affective Neuroscience*, *7*(6), 738–742. https://doi.org/10.1093/scan/nss059
Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., … Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*, 21. https://doi.org/10.1038/s41562-016-0021
Neubauer, A. C., & Fink, A. (2009). Intelligence and neural efficiency. *Neuroscience &*

Biobehavioral Reviews, *33*(7), 1004–1023. https://doi.org/10.1016/J.NEUBIOREV.2009.04.001

Neuroskeptic. (2012). The nine circles of scientific hell. *Perspectives on Psychological Science*, *7*(6), 643–644. https://doi.org/10.1177/1745691612459519

Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Human Brain Mapping*, *15*(1), 1–25. https://doi.org/10.1002/hbm.1058

Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E.-J. (2011). Erroneous analyses of interactions in neuroscience: a problem of significance. *Nature Neuroscience*, *14*(9), 1105–1107. https://doi.org/10.1038/nn.2886

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., … Huettig, F. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *ELife*, *7*. https://doi.org/10.7554/eLife.33468

Nissen, S. B., Magidson, T., Gross, K., & Bergstrom, C. T. (2016). Publication bias and the canonization of false facts. *ELife*, *5*. https://doi.org/10.7554/eLife.21451

Nord, C. L., Valton, V., Wood, J., & Roiser, J. P. (2017). Power-up: a reanalysis of "power failure" in neuroscience using mixture modeling. *The Journal of Neuroscience*, *37*(34), 8051–8061. https://doi.org/10.1523/JNEUROSCI.3592-16.2017

Oechslin, M. S., Meyer, M., & Jäncke, L. (2010). Absolute pitch — functional evidence of speech-relevant auditory acuity. *Cerebral Cortex*, *20*(2), 447–455. https://doi.org/10.1093/cercor/bhp113

Ohnishi, T., Matsuda, H., Asada, T., Aruga, M., Hirakata, M., Nishikawa, M., … Imabayashi, E. (2001). Functional anatomy of musical perception in musicians. *Cerebral Cortex*, *11*(8), 754–760. https://doi.org/10.1093/cercor/11.8.754

Open Science Collaboration, O. S. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Oxenham, A. J. (2012). Pitch perception. *The Journal of Neuroscience*, *32*(39), 13335–13338. https://doi.org/10.1523/JNEUROSCI.3815-12.2012

Oxenham, A. J. (2018). How We Hear: The Perception and Neural Coding of Sound. *Ann. Rev. of Psychol.*, *69*(1), 27–50. https://doi.org/10.1146/annurev-psych-122216-011635

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature*, *392*(6678), 811–814. Retrieved from http://dx.doi.org/10.1038/33918

Pascual-Marqui, R. D., Sekihara, K., Brandeis, D., & Michel, C. M. (2009). Imaging the electric neuronal generators of EEG/MEG. In C. M. Michel, T. Koenig, D. Brandeis, L. R. R. Gianotti, & J. Wackermann (Eds.), *Electrical neuroimaging* (pp. 49–78). Cambridge University Press. https://doi.org/10.1017/CBO9780511596889.004

Patel, A. D. (2007). *Music, Language, and the Brain*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195123753.001.0001

Peretz, I. (2016). Neurobiology of Congenital Amusia. *Trends in Cognitive Sciences*, *20*(11), 857–867. https://doi.org/10.1016/J.TICS.2016.09.002

Petrides, M., Alivisatos, B., Evans, A. C., & Meyer, E. (1993). Dissociation of human mid-dorsolateral from posterior dorsolateral frontal cortex in memory processing. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(3), 873–877. https://doi.org/10.1073/pnas.90.3.873

Plack, C. J., Barker, D., & Hall, D. A. (2014). Pitch coding and pitch processing in the human brain. *Hearing Research*, *307*, 53–64. https://doi.org/10.1016/j.heares.2013.07.020

Plack, C. J., Oxenham, A. J., Fay, R. R., & Popper, A. N. (Eds.). (2005). *Pitch: Neural Coding and Perception.* Springer.

Plantinga, J., & Trainor, L. J. (2005). Memory for melody: Infants use a relative pitch code. *Cognition*, *98*(1), 1–11. https://doi.org/10.1016/J.COGNITION.2004.09.008

Poldrack, R. A. (2007). Region of interest analysis for fMRI. *Social Cognitive and Affective Neuroscience*, *2*(1), 67–70. https://doi.org/10.1093/scan/nsm006

Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R.,

… Yarkoni, T. (2017). Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience*, *18*(2), 115–126. https://doi.org/10.1038/nrn.2016.167

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Profita, J., & Bidder, T. G. (1988). Perfect pitch. *American Journal of Medical Genetics*, *29*(4), 763–771. https://doi.org/10.1002/ajmg.1320290405

Ratcliff, R., Philiastides, M. G., & Sajda, P. (2009). Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(16), 6539–6544. https://doi.org/10.1073/pnas.0812589106

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*(6), 718–724. https://doi.org/10.1038/nn.2331

Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(22), 11800–11806. https://doi.org/10.1073/pnas.97.22.11800

Rogenmoser, L., Elmer, S., & Jäncke, L. (2015). Absolute pitch: evidence for early cognitive facilitation during passive listening as revealed by reduced P3a amplitudes. *Journal of Cognitive Neuroscience*, *27*(3), 623–637. https://doi.org/10.1162/jocn_a_00708

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Ross, D. A., Gore, J. C., & Marks, L. E. (2005). Absolute pitch: Music and beyond. *Epilepsy & Behavior*, *7*(4), 578–601. https://doi.org/10.1016/j.yebeh.2005.05.019

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374. https://doi.org/10.1016/J.JMP.2012.08.001

Rouder, J. N., Morey, R. D., Verhagen, J., Swagman, A. R., & Wagenmakers, E.-J. (2017). Bayesian analysis of factorial designs. *Psychological Methods*, *22*(2), 304–321. https://doi.org/10.1037/met0000057

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Sacks, O. (1995). Musical ability. *Science*, *268*(5211), 621–622. https://doi.org/10.1126/science.7732360

Sankaran, N., Swaminathan, J., Micheyl, C., Kalluri, S., & Carlile, S. (2018). Tracking the dynamic representation of consonants from auditory periphery to cortex. *The Journal of the Acoustical Society of America*, *144*(4), 2462–2472. https://doi.org/10.1121/1.5065492

Schaefer, R. S., Farquhar, J., Blokland, Y., Sadakata, M., & Desain, P. (2011). Name that tune: Decoding music from the listening brain. *NeuroImage*, *56*(2), 843–849. https://doi.org/10.1016/J.NEUROIMAGE.2010.05.084

Schellenberg, E. G., & Trehub, S. E. (2003). Good Pitch Memory Is Widespread. *Psychological Science*, *14*(3), 262–266. https://doi.org/10.1111/1467-9280.03432

Schlaug, G., Jäncke, L., Huang, Y., & Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in musicians. *Science*, *267*(5198), 699–701. https://doi.org/10.1126/science.7839149

Schlemmer, K. B., Kulke, F., Kuchinke, L., & Van Der Meer, E. (2005). Absolute pitch and pupillary response: Effects of timbre and key color. *Psychophysiology*, *42*(4), 465–472. https://doi.org/10.1111/j.1469-8986.2005.00306.x

Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., & Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nature Neuroscience*, *5*(7), 688–694. https://doi.org/10.1038/nn871

Schneider, P., Sluming, V., Roberts, N., Scherg, M., Goebel, R., Specht, H. J., … Rupp, A. (2005). Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. *Nature Neuroscience*, *8*(9), 1241–1247.

https://doi.org/10.1038/nn1530
Schulze, K., Gaab, N., & Schlaug, G. (2009). Perceiving pitch absolutely: comparing absolute and relative pitch possessors in a pitch memory task. *BMC Neuroscience*, *10*(1), 1–13. https://doi.org/10.1186/1471-2202-10-106
Schulze, K., Koelsch, S., & Williamson, V. (2018). Auditory working memory. In R. Bader (Ed.), *Springer Handbook of Systematic Musicology* (pp. 461–472). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-662-55004-5_24
Schulze, K., Mueller, K., & Koelsch, S. (2013). Auditory stroop and absolute pitch: an fMRI study. *Human Brain Mapping*, *34*(7), 1579–1590. https://doi.org/10.1002/hbm.22010
Shah, N. J., Steinhoff, S., Mirzazade, S., Zafiris, O., Grosse-Ruyken, M. L., Jäncke, L., & Zilles, K. (2000). The effect of sequence repeat time on auditory cortex stimulation during phonetic discrimination. *NeuroImage*, *12*(1), 100–108. https://doi.org/10.1006/nimg.2000.0588
Siegel, J. A. (1974). Sensory and verbal coding strategies in subjects with absolute pitch. *Journal of Experimental Psychology*, *103*(1), 37–44. https://doi.org/10.1037/h0036844
Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632
Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, *9*(1), 76–80. https://doi.org/10.1177/1745691613514755
Smith, N. A., & Schmuckler, M. A. (2008). Dial A440 for absolute pitch: Absolute pitch memory by non-absolute pitch possessors. *The Journal of the Acoustical Society of America*, *123*(4), EL77–EL84. https://doi.org/10.1121/1.2896106
Steblin, R. (1987). Towards a History of Absolute Pitch Recognition. *College Music Symposium*, *27*, 141–153. Retrieved from https://www.jstor.org/stable/pdf/40373845.pdf?refreqid=excelsior%3A9c364ac5d67280ab3d346791479ef3b4
Stelzer, J., Chen, Y., & Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): Random permutations and cluster size control. *NeuroImage*, *65*, 69–82. https://doi.org/10.1016/j.neuroimage.2012.09.063
Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651
Supekar, K., Uddin, L. Q., Khouzam, A., Phillips, J., Gaillard, W. D., Kenworthy, L. E., … Menon, V. (2013). Brain Hyperconnectivity in Children with Autism and its Links to Social Deficits. *Cell Reports*, *5*(3), 738–747. https://doi.org/10.1016/J.CELREP.2013.10.001
Szucs, D., & Ioannidis, J. P. A. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLOS Biology*, *15*(3), e2000797. https://doi.org/10.1371/journal.pbio.2000797
Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Brainstorm: A user-friendly application for MEG/EEG analysis. *Computational Intelligence and Neuroscience*, *2011*, 13. https://doi.org/10.1155/2011/879716
Takeuchi, A. H., & Hulse, S. H. (1991). Absolute-pitch judgments of black and white-key pitches. *Music Perception: An Interdisciplinary Journal*, *9*(1), 27–46. https://doi.org/10.2307/40286157
Takeuchi, A. H., & Hulse, S. H. (1993). Absolute pitch. *Psychological Bulletin*, *113*(2), 345–361. https://doi.org/10.1037/0033-2909.113.2.345
Terhardt, E., & Seewann, M. (1983). Aural Key Identification and Its Relationship to Absolute Pitch. *Music Perception: An Interdisciplinary Journal*, *1*(1), 63–83. https://doi.org/10.2307/40285250
Terhardt, E., & Ward, W. D. (1982). Recognition of musical key: Exploratory study. *The Journal of the Acoustical Society of America*, *72*(1), 26–33. https://doi.org/10.1121/1.387989
Tervaniemi, M., Alho, K., Paavilainen, P., Sams, M., & Näätänen, R. (1993). Absolute pitch and event-related brain potentials. *Music Perception: An Interdisciplinary Journal*, *10*(3),

305. https://doi.org/10.2307/40285572

Theusch, E., Basu, A., & Gitschier, J. (2009). Genome-wide study of families with absolute pitch reveals linkage to 8q24.21 and locus heterogeneity. *The American Journal of Human Genetics*, *85*(1), 112–119. https://doi.org/10.1016/j.ajhg.2009.06.010

Theusch, E., & Gitschier, J. (2011). Absolute pitch twin study and segregation analysis. *Twin Research and Human Genetics*, *14*(2), 173–178. https://doi.org/10.1375/twin.14.2.173

Uddin, L. Q., Supekar, K., Lynch, C. J., Cheng, K. M., Odriozola, P., Barth, M. E., … Menon, V. (2015). Brain State Differentiation and Behavioral Inflexibility in Autism. *Cerebral Cortex*, *25*(12), 4740–4747. https://doi.org/10.1093/cercor/bhu161

Van Hedger, S. C., Heald, S. L. M., Huang, A., Rutstein, B., & Nusbaum, H. C. (2017). Telling in-tune from out-of-tune: widespread evidence for implicit absolute intonation. *Psychonomic Bulletin & Review*, *24*(2), 481–488. https://doi.org/10.3758/s13423-016-1099-1

Van Hedger, S. C., Heald, S. L. M., Koch, R., & Nusbaum, H. C. (2015). Auditory working memory predicts individual differences in absolute pitch learning. *Cognition*, *140*, 95–110. https://doi.org/10.1016/J.COGNITION.2015.03.012

Van Hedger, S. C., Heald, S. L. M., & Nusbaum, H. C. (2015). The effects of acoustic variability on absolute pitch categorization: Evidence of contextual tuning. *The Journal of the Acoustical Society of America*, *138*(1), 436–446. https://doi.org/10.1121/1.4922952

Van Hedger, S. C., Heald, S. L. M., & Nusbaum, H. C. (2016). What the [bleep]? Enhanced absolute pitch memory for a 1000 Hz sine tone. *Cognition*, *154*, 139–150. https://doi.org/10.1016/j.cognition.2016.06.001

Van Hedger, S. C., Heald, S. L. M., & Nusbaum, H. C. (2018a). Absolute pitch can be learned by some adults. *BioRxiv*, 325050. https://doi.org/10.1101/325050

Van Hedger, S. C., Heald, S. L. M., Uddin, S., & Nusbaum, H. C. (2018). A note by any other name: Intonation context rapidly changes absolute note judgments. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(8), 1268–1282. https://doi.org/10.1037/xhp0000536

Van Hedger, S. C., Heald, S. L., & Nusbaum, H. C. (2018b). Long-term pitch memory for music recordings is related to auditory working memory precision. *Quarterly Journal of Experimental Psychology*, *71*(4), 879–891. https://doi.org/10.1080/17470218.2017.1307427

Van Hedger, S. C., & Nusbaum, H. C. (2018). Individual differences in absolute pitch performance: Contributions of working memory, musical expertise, and tonal language background. *Acta Psychologica*, *191*, 251–260. https://doi.org/10.1016/J.ACTPSY.2018.10.007

Vanzella, P., & Schellenberg, E. G. (2010). Absolute pitch: Effects of timbre on note-naming ability. *PLoS ONE*, *5*(11), e15449. https://doi.org/10.1371/journal.pone.0015449

Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science*, *4*(3), 274–290. https://doi.org/10.1111/j.1745-6924.2009.01125.x

Wang, X.-J. (2010). Neurophysiological and Computational Principles of Cortical Rhythms in Cognition. *Physiological Reviews*, *90*(3), 1195–1268. https://doi.org/10.1152/physrev.00035.2008

Wang, X., & Walker, K. M. M. (2012). Neural Mechanisms for the Abstraction and Use of Pitch Information in Auditory Cortex. *The Journal of Neuroscience*, *32*(39), 13339–13342. https://doi.org/10.1523/JNEUROSCI.3814-12.2012

Ward, W. D. (1999). Absolute Pitch. In *The Psychology of Music* (pp. 265–298). Academic Press. https://doi.org/10.1016/B978-012213564-4/50009-3

Ward, W. D., & Burns, E. M. (1982). Absolute Pitch. In *The Psychology of Music* (pp. 431–451). Elsevier. https://doi.org/10.1016/B978-0-12-213562-0.50018-8

Wayman, J. W., Frisina, R. D., Walton, J. P., Hantz, E. C., & Crummer, G. C. (1992). Effects of musical training and absolute pitch ability on event-related activity in response to sine tones. *The Journal of the Acoustical Society of America*, *91*(6), 3527–3531. https://doi.org/10.1121/1.402841

Wengenroth, M., Blatow, M., Heinecke, A., Reinhardt, J., Stippich, C., Hofmann, E., & Schneider, P. (2014). Increased volume and function of right auditory cortex as a marker for absolute pitch. *Cerebral Cortex*, *24*(5), 1127–1137. https://doi.org/10.1093/cercor/bhs391

Wenhart, T., & Altenmüller, E. (2019). A Tendency Towards Details? Inconsistent Results on Auditory and Visual Local-To-Global Processing in Absolute Pitch Musicians. *Frontiers in Psychology*, *10*, 31. https://doi.org/10.3389/fpsyg.2019.00031

Wenhart, T., Bethlehem, R. A. I., Baron-Cohen, S., & Altenmüller, E. (2019). Autistic traits, resting-state connectivity, and absolute pitch in professional musicians: shared and distinct neural features. *Molecular Autism*, *10*(1), 20. https://doi.org/10.1186/s13229-019-0272-6

Wenhart, T., Hwang, Y.-Y., & Altenmüller, E. (2019). Enhanced auditory disembedding in an interleaved melody recognition test is associated with absolute pitch ability. *Scientific Reports*, *9*(1), 7838. https://doi.org/10.1038/s41598-019-44297-x

West Marvin, E., VanderStel, J., & Siu, J. C.-S. (2019). In their own words: Analyzing the extents and origins of absolute pitch. *Psychology of Music*, 030573561983295. https://doi.org/10.1177/0305735619832959

Whitham, E. M., Pope, K. J., Fitzgibbon, S. P., Lewis, T., Clark, C. R., Loveless, S., … Willoughby, J. O. (2007). Scalp electrical recording during paralysis: Quantitative evidence that EEG frequencies above 20 Hz are contaminated by EMG. *Clinical Neurophysiology*, *118*(8), 1877–1888. https://doi.org/10.1016/J.CLINPH.2007.04.027

Wilson, S. J., Lusher, D., Wan, C. Y., Dudgeon, P., & Reutens, D. C. (2009). The neurocognitive components of pitch processing: Insights from absolute pitch. *Cerebral Cortex*, *19*(3), 724–732. https://doi.org/10.1093/cercor/bhn121

Wilson, S. J., Martin, C. L., Martin, C. L., Rayner, G., & McLachlan, N. (2012). Intersecting Factors Lead to Absolute Pitch Acquisition That is Maintained in a "Fixed do " Environment. *Music Perception: An Interdisciplinary Journal*, *29*(3), 285–296. https://doi.org/10.1525/mp.2012.29.3.285

Wise, R. J. S., Scott, S. K., Blank, S. C., Mummery, C. J., Murphy, K., & Warburton, E. A. (2001). Separate neural subsystems within `Wernicke's area'. *Brain*, *124*(1), 83–95. https://doi.org/10.1093/brain/124.1.83

Wong, Y. K., Lui, K. F. H., Yip, K. H. M., & Wong, A. C.-N. (2019). Is it impossible to acquire absolute pitch in adulthood? *BioRxiv*, 355933. https://doi.org/10.1101/355933

Zatorre, R. J. (2003). Absolute pitch: A model for understanding the influence of genes and development on neural and cognitive function. *Nature Neuroscience*, *6*(7), 692–695. https://doi.org/10.1038/nn1085

Zatorre, R. J., & Beckett, C. (1989). Multiple coding strategies in the retention of musical tones by possessors of absolute pitch. *Memory & Cognition*, *17*(5), 582–589. https://doi.org/10.3758/BF03197081

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences*, *6*(1), 37–46. https://doi.org/10.1016/S1364-6613(00)01816-7

Zatorre, R. J., Evans, A. C., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *The Journal of Neuroscience*, *14*(4), 1908. https://doi.org/10.1523/JNEUROSCI.14-04-01908.1994

Zatorre, R. J., Perry, D. W., Beckett, C. A., Westbury, C. F., & Evans, A. C. (1998). Functional anatomy of musical processing in listeners with absolute pitch and relative pitch. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(6), 3172–3177. https://doi.org/10.1073/pnas.95.6.3172

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, *41*, e120. https://doi.org/10.1017/S0140525X17001972